

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:35:16 ; Search time 13.43 Seconds
(without alignments)
980.242 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MRLGILLOGLTRFHTSPPTD.....CARGRVETNNEGEPVSVMF 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107.5	5.9	1533	1 PUM_DROME	P25822 drosophila
2	105	5.8	480	1 TGN5_HUMAN	Q43493 homo sapien
3	102	5.6	1914	1 KML5_HUMAN	Q15746 homo sapien
4	97	5.3	980	1 BOB1_YEAST	P38041 saccharomyc
5	96	5.3	367	1 CSP_PLAYO	P06914 plasmodium
6	96	5.3	892	1 ATX7_HUMAN	Q15265 homo sapien
7	95	5.2	1168	1 MYSC_ACACA	P10569 acanthameob
8	94.5	5.2	509	1 SOX9_PIG	Q18896 sus scrofa
9	94.5	5.2	602	1 CRK_DAUCA	P53681 daucus caro
10	93	5.1	507	1 MEFA_HUMAN	Q02078 homo sapien
11	93	5.1	509	1 SOX9_HUMAN	P48436 homo sapien
12	93	5.1	579	1 SYN3_RAT	O70441 rattus norv
13	93	5.1	604	1 RYOL_NEUCR	P78706 neurospora
14	93	5.1	1329	1 FTSK_ECOLI	P46889 escherichia
15	92.5	5.1	498	1 MEFA_MOUSE	Q06929 mus musculu
16	92	5.1	363	1 TOBI_MOUSE	Q61471 mus musculu
17	91.5	5.0	411	1 Y513_HUMAN	O60268 homo sapien
18	91.5	5.0	785	1 SOK2_YEAST	P05248 saccharomyc
19	90.5	5.0	877	1 WEL1_SCHPO	P07527 schizosacch
20	90.5	5.0	1047	1 HIRA_DROME	O17468 drosophila
21	90.5	5.0	2515	1 TUD_DROME	P25823 drosophila
22	89.5	4.9	1460	1 PMPC_CHLMU	Q9pjvl chlamydia m
23	89	4.9	673	1 ESUN2_MICUN	P57781 micropogoni
24	88.5	4.9	393	1 SMA3_CASEL	P45896 caenorhabdi
25	88.5	4.9	649	1 INVA_PHAAU	P29001 phaseolus a
26	88.5	4.9	1262	1 CA13_CHICK	P12105 gallus gall
27	88	4.8	345	1 TOBI_HUMAN	P50616 homo sapien
28	88	4.8	734	1 YHU9_YEAST	P32900 saccharomyc
29	88	4.8	838	1 GLT4_WHEAT	P08489 triticum ae
30	87.5	4.8	580	1 SYN3_HUMAN	O14994 homo sapien
31	87.5	4.8	1527	1 CAIH_MOUSE	P39061 mus musculu
32	87	4.8	304	1 UL49_HSVEB	P28960 equine herp
33	87	4.8	609	1 RFA1_XENLA	Q01588 xenopus lae

RESULT 1

ID	PUM_DROME	STANDARD;	PRT;	1533 AA.
AC	P25822;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Maternal pumilio protein.			
GN	PUM.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=92249205; Pubmed=1576962;			
RA	Macdonald P.M.;			
RT	"The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein.";			
RT	Development 114:221-234(1992).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=93093466; Pubmed=1459455;			
RA	Barker D.D., Wang C., Moore J., Dickinson L.K., Lehmann R.;			
RT	*Pumilio is essential for function but not for distribution of the Drosophila abdominal determinant Nanos.*;			
RL	Genes Dev. 6:2312-2326(1992).			
CC	-!- FUNCTION: PUM IS THE ONLY GENE REQUIRED FOR NOS ACTIVITY THAT IS NOT ALSO REQUIRED FOR POSTERIOR LOCALIZATION OF GERM LINE DETERMINANTS. PUM IS REQUIRED DURING EMBRYOGENESIS WHEN NOS ACTIVITY APPARENTLY MOVES ANTERIORLY FROM THE POSTERIOR POLE.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC. IT IS CONCENTRATED IN THE CORTICAL REGION OF THE EMBRYO BENEATH THE NUCLEI.			
CC	-!- DEVELOPMENTAL STAGE: EXPRESSED IN THE OVARIES AND DURING THE EMBRYOGENESIS.			
CC	-!- DOMAIN: CONSISTS MAINLY OF REGIONS ENRICHED IN A SINGLE AMINO ACID.			
CC	-!- DISEASE: LETHAL DEFECTIVE IN POSTERIOR PATTERN FORMATION.			
CC	-!- SIMILARITY: BELONGS TO THE PUMILIO/MPT5 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X62589; CAA44474.1; -			
DR	EMBL; L07943; AAB59189.1; -			
DR	PIR; S22026; S22026.			
DR	HSSP; P04002; LWPA.			
DR	Flybase; FBgn0003165; pum.			
DR	InterPro; IPR001313; PUM.			
DR	Pfam; PF00806; PUF; 8.			

ALIGNMENTS

34	86.5	4.8	521	1	VGLC_HSVBC	PI4378 bovine herp
35	86.5	4.8	651	1	INVA_PHAVU	O24509 phaseolus v
36	86.5	4.8	1461	1	IE18_PRIVF	PI1675 pseudorabie
37	86	4.7	805	1	TAC1_HUMAN	O75410 homo sapien
38	86	4.7	1490	1	CRK7_HUMAN	Q9nyv4 homo sapien
39	85.5	4.7	254	1	RCEH_RHOCA	PI9056 rhodobacter
40	85.5	4.7	1051	1	TF1A_MOUSE	Q64127 mus musculu
41	85.5	4.7	1516	1	CA1H_HUMAN	P39060 homo sapien
42	85	4.7	574	1	LOL1_HUMAN	Q08397 homo sapien
43	85	4.7	630	1	MLA3_HUMAN	Q08r34 homo sapien
44	85	4.7	1192	1	RTN4_HUMAN	Q9nqc3 homo sapien
45	84.5	4.6	1584	1	BAIL_HUMAN	O14514 homo sapien


```
FT DOMAIN 139 228 15 X 6 AA TANDEM REPEATS OF Q-G-P-G-A-P.
FT DOMAIN 229 260 8 X 4 AA TANDEM REPEATS OF Q-Q-P-P.
SQ SEQUENCE 367 AA: 38888 MW: 1EA56AFF7FECB5E3 CRC64;

Query Match 5.3%; Score 96; DB 1; Length 367;
Best Local Similarity 38.1%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 31; Indels 12; Gaps 5;

QY 72 PRAQPLPAQPPALAPQVQSPQPPQTRVWAPRNNAAFQSGGAGSDNSNPGVQPNRS- 130
Db 231 PQQP-PQPP--QQPPQPPQPPQPPQPPQPPR-----PPDGNNNNNNNNGNNSDY 281
QY 131 APSVESHPVLEKKAHSYNPKF 154
Db 282 VPSAEQ--ILEFVKQISSOLTEEM 303

RESULT 6
ATX7_HUMAN STANDARD; PRT; 892 AA.
AC O15265; O75328; Q9Y6P8;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin 7 (Spinocerebellar ataxia type 7 protein).
GN SCA7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RX MEDLINE=97434213; PubMed=9288099;
RA David G., Abbas N., Stavanin G., Duerr A., Yvert G., Cancel G.,
RA Weber C., Imbert G., Saudou F., Antoniou E., Drabkin H., Gemmill R.,
RA Giunti P., Benomar A., Wood N., Ruberg M., Agid Y., Mandel J.-L.,
RA Brice A.;
RT "Cloning of the SCA7 gene reveals a highly unstable CAG repeat
RT expansion.";
RL Nat. Genet. 17:65-70(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=98087568; PubMed=9425224;
RA Del-Favero J., Krols L., Michalik A., Theuns J., Loeferen A.,
RA Goossens D., Wehnert A., Van den Bossche D., Van Zand K.,
RA Backhovens H., van Regenmortel N., Martin J.-J., Van Broeckhoven C.;
RT "Molecular genetic analysis of autosomal dominant cerebellar ataxia
RT with retinal degeneration (ADCA type II) caused by CAG triplet repeat
RT expansion.";
RL Hum. Mol. Genet. 7:177-186(1998).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99371772; PubMed=1041328;
RA Kaytor M.D., Duvick L.A., Skinner P.J., Koob M.D., Ranum L.P.,
RA Orr H.T.;
RT "Nuclear localization of the spinocerebellar ataxia type 7 protein,
RT ataxin-7.";
RL Hum. Mol. Genet. 8:1657-1664(1999).
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN ADDITION TO A DIFFUSE
CC DISTRIBUTION THROUGHOUT THE NUCLEUS, IT IS ASSOCIATED WITH THE
CC NUCLEAR MATRIX AND THE NUCLEOLUS.
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF SCA7 IS HIGHLY POLYMORPHIC (4
CC TO 18 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT
CC 38-130 REPEATS IN SCA7 PATIENTS. INTERMEDIATE ALLELES WITH 28 TO
CC 35 REPEATS ARE PRONE TO FURTHER EXPANSION.
CC -!- DISEASE: DEFECTS IN SCA7 ARE THE CAUSE OF SPINOCEREBELLAR ATAXIA
CC TYPE 7; ALSO KNOWN AS AUTOSOMAL DOMINANT CEREBELLAR ATAXIA TYPE II
CC (ADCA-II). ADCA-II IS CHARACTERIZED BY PIGMENTARY MACULAR
CC DYSTROPHY AND RETINAL DEGENERATION LEADING TO BLINDNESS.
CC -!- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ000517; CAA04154.1; -
DR EMBL: AF032105; AAC39765.1; -
DR EMBL: AF032102; AAC19162.1; -
DR EMBL: AF032103; AAC19163.1; -
DR MIM: 164500; -
DR InterPro: IPR000822; Znf-C2H2.
DR SMART: SM00355; Znf_C2H2; 1.
KW Polymorphism; Triplet repeat expansion; Disease mutation;
KW Nuclear protein.
FT DOMAIN 16 20 POLY-ALA.
FT DOMAIN 23 28 POLY-ALA.
FT DOMAIN 30 39 POLY-GLN.
FT DOMAIN 40 45 POLY-PRO.
FT DOMAIN 51 55 POLY-PRO.
FT DOMAIN 171 174 POLY-SER.
FT DOMAIN 213 219 POLY-SER.
FT DOMAIN 647 654 POLY-SER.
FT DOMAIN 717 730 POLY-SER.
FT DOMAIN 840 845 POLY-SER.
FT VARIANT 284 264 K -> R (IN DBSNP:1053338).
FT VARIANT 663 /FTIG-VAR_011823.
FT VARIANT 663 P -> S (IN DBSNP:1053340).
FT CONFLICT 105 105 P -> H (IN REF. 2; AAC19162).
FT CONFLICT 129 129 C -> S (IN REF. 2; AAC19162).
FT CONFLICT 862 862 V -> M (IN REF. 2; AAC19162).
FT CONFLICT 888 PKARP -> VGNGL (IN REF. 2; AAC39765 AND
FT AF032103).
SQ SEQUENCE 892 AA; 95451 MW; 9AEA787D77103C5F CRC64;

Query Match 5.3%; Score 96; DB 1; Length 892;
Best Local Similarity 20.1%; Pred. No. 10;
Matches 73; Conservative 45; Mismatches 124; Indels 122; Gaps 16;

QY 17 PPTDSSVTETI-----ILCTMLFLGSLGAWGTTSTIGSIFSLKTLRSHQGVGL 67
Db 567 PSTTSPITRIPHTNSVPTSCQGVSYLAA-----ATVSTSPVLLSSTCIS----- 612
QY 68 KYSRPAQPLPAQPPAL-AQPYQSPQPPQTRVWAPRNNAAFQSGGAGSDNSNPGV 126
Db 613 ----PNSKSVPAHGTTLNAQPAASGAMD-----VCSMQSRQVSSSSSPST- 655
QY 127 QPNAPSVEHPVL---EKLKAHSYNPKFEWN-----LKSGRVFIKSYSEDD 173
Db 656 -PSGLSSVPSPMRKPQKLKSKSLRPKESNGSTNCQNASSTSGSGGKKRKNSSPL 714
QY 174 IHRSTIKYITWCSTEGNKRDLDSAFR--CMSSGPGVYLLFSVNGSGHFCGVAEMKSPDYG 231
Db 715 VHSSSSSSSSSSSSSSHME---SFRKNCVAHSGPPY--STVTSSHSIGL----- 758
QY 232 TSAGVNSQDKWKGFQVQWIFVKDVPNNQLR--HRLNNDNKPVTNSRDTQEVPLEKAK 289
Db 759 -----NCVTNKANAVNVRHDSGRGPPGTS-----PAESIK 789
QY 290 OVLKIISYKHTTSI-----FDDFAH-----YEKRQRRRRCARNRVETNN 331
Db 790 RSMVWNSSDSTLSLGPFTIHQSNELPVNSHGFSFSSHTPLDLKLGKRRKCSPPSSSINN 849
QY 332 ECEP 335
Db 850 SSKP 853
```

```

RESULT 7
MYSC_ACACA STANDARD; PRT; 1168 AA.
AC P10569;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MIC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RT and non-myosin-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86259656; PubMed=3014500;
RA Hammer J.A. III, Jung G., Korn E.D.;
RT "Genetic evidence that Acanthamoeba myosin I is a true myosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
RN [3]
RP PHOSPHORYLATION SITE.
RX MEDLINE=90037074; PubMed=2530230;
RA Bzjeska H., Lynch T.J., Martin B., Korn E.D.;
RT "The localization and sequence of the phosphorylation sites of
RT Acanthamoeba myosin I. An improved method for locating the
RT phosphorylated amino acid.";
RL J. Biol. Chem. 264:19340-19348(1989).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
CC OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -1- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J02974; AAA27707.1; -.
CC PIR: A33891; MWAXIC.
CC HSP: P08799; ILVK.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00242; MYSC; 1.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS50002; SH3; 1.
CC MYosin: ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
FT DOMAIN 671 922 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).

```

```

FT DOMAIN 976 1035 SH3.
FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
FT NP_BIND 101 108 ATP (POTENTIAL).
FT MOD_RES 311 311 PHOSPHORYLATION.
SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;

Query Match 5.2%; Score 95; DB 1; Length 1168;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 41; Conservative 20; Mismatches 68; Indels 56; Gaps 7;

QY 45 SISTGSIFSILKTRSQHGQVGLKVSRAQPLPAQPALAQQYQSPQQ----- 94
Db 857 TLGGKTCVVKFIRDPOGDG--KVKGTQSVAPGLPSSA-PNIAQPETSGGASFTVA 913
QY 95 -----PPQTRWVAPRNRAAFQSGGAGSDNSPQNVQPN 130
Db 914 EQSYKQDILGAKGGGGGRGSPGSPGAVSRP-----SPGGGGGSPFGGRPSG 968
QY 131 APSVESHVLEKKAHSY---NPKFEWNLKSRVFIKSYSEDDIHRSIKYSIWGSTE 187
Db 969 PPAASAPGPEQARALYDFAENDELTFN--EGAVTVINKSNPD-----WHEGE 1017
QY 188 HGKRR 192
Db 1018 LMGQR 1022

RESULT 8
SOX9_PIG STANDARD; PRT; 509 AA.
AC O18836;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription factor SOX-9.
GN SOX9.
OS Sus scrofa (Pig).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Pilon N., Silversides D.W.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE NORMAL SKELETAL
CC DEVELOPMENT. MAY REGULATE THE EXPRESSION OF OTHER GENES INVOLVED
CC IN CHONDROGENESIS BY ACTING AS A TRANSCRIPTION FACTOR FOR THESE
CC GENES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF029696; AAB81431.1; -.
CC HSP: P48436; ISX9.
CC InterPro: IPR000910; HMG_12_box.
CC Pfam: PF00505; HMG_box; 1.
CC SMART: SM00398; HMG; 1.
CC Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 105 173 HMG BOX.
FT DOMAIN 339 378 GLN/PRO-RICH.
FT DOMAIN 342 349 POLY-PRO.
SQ SEQUENCE 509 AA; 56267 MW; 069F77A0510BBFC2 CRC64;

Query Match 5.2%; Score 94.5; DB 1; Length 509;

```

FT	REPEAT	20	26	1.
FT	REPEAT	27	33	2.
FT	REPEAT	34	40	3.
FT	DOMAIN	148	410	PROTEIN KINASE.
FT	NP_BIND	154	162	ATP (BY SIMILARITY).
FT	BINDING	180	180	ATP (BY SIMILARITY)..
FT	ACT_SITE	276	276	BY SIMILARITY.
FT	DOMAIN	465	476	ANCESTRAL CALCIUM SITE 1 (POTENTIAL).
FT	DOMAIN	502	513	ANCESTRAL CALCIUM SITE 2 (POTENTIAL).
FT	DOMAIN	542	553	ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT	DOMAIN	574	585	ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
SQ	SEQUENCE	602 AA; 67184 MW; 1D10BF68B37BF447 CRC64;		

Query Match 5.2%; Score 94.5; DB 1; Length 602;
Best Local Similarity 19.9%; Pred. No. 8.1;
Matches 56; Conservative 39; Mismatches 86; Indels 101; Gaps 12;

QY	65	VGLKVSRRAPQ-----LPAQ-----PPALAQPYCGSPQP----	POTRW 100
	:	: : : :	: : :
Db	1	MGICVSKSPPEPLDHNHHTSIPVNDTSLPDODNSIPPKDIAIPAQTNNKPKKKSPFLPF 60	
QY	101	VAPRNRAAFCQ-----SCGASDSNPGNVQNAPSVESHPVEKLKAASHYNPKEF 154	
	:	: : : :	: : :
Db	61	YSPAHFLEFSKSPAVGVSPAAGSSNSTPKRLFPFPPSPAKH-----IKAA----- 107	
QY	155	EWNLKSGRV---FIIKSYSEDTHRSIKYSIMCSTEHGNKRLDSPACRCHSGPVLLF 210	
	:	: : : :	: : :
Db	108	-WARRGSVPNCAAPENNEVD-----GGALDKSFSGFKFGSKFEVG 151	
QY	211	SVANGSGHFCGVAEMKSPVDYGTSGAVMWSQKW-KGKFEDVQWIFVKYPNNQLRHIRENN 269	
	:	: : : :	: : :
Db	152	EEVGRGHF-----GYTCRAKFKGEFGQDVAVKVIPKAKM----- 187	
QY	270	DNKPVTNSRDTOEVPLEKAK---QVLKIISYKHHTSIFDDF 308	
	:	: : : :	: : :
Db	188	-----TTAIAIEDVRREVVKILRALTGHHNLVQFYDAF 219	

RESULT 10
MEFA_HUMAN

ID	MEFA_HUMAN	STANDARD;	PRT;	507 AA.
QC	Q02078; Q14223; Q14224; Q43814;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Myocyte-specific enhancer factor 2A (Serum response factor-like protein 1).			
GN	MEF2A OR MEF2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Heart, and Skeletal muscle;			
RX	MEDLINE=92387551; PubMed=1516833;			
RA	Yu Y.-T., Breitbar R.E., Smoot L.B., Lee Y., Mahdavi V.,			
RA	Nadal-Ginard B.;			
RT	"Human myocyte-specific enhancer factor 2 comprises a group of			
RT	tissue-restricted MADS box transcription factors.";			
RL	Genes Dev. 6:1783-1798(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=92084105; PubMed=1748287;			
RA	Pollock R., Treisman R.;			
RT	"Human SRF-related proteins: DNA-binding properties and potential			
RT	regulatory targets.";			
RL	Genes Dev. 5:2327-2341(1991).			
RN	[3]			
RP	REVISIONS, AND SEQUENCE FROM N.A. (ISOFORM RSFRCS).			
RA	Treisman R.;			

RT autosomal sex reversal.";
 RL Am. J. Hum. Genet. 57:1028-1036(1995).
 RN [5]
 RP VARIANTS CD LEU-108; ARG-143; PRO-152 AND ARG-170.
 RX MEDLINE=97156215; PubMed=9002675;
 RA Meyer J., Suedbeck P., Held M., Wagner T., Schmitz M.L.,
 RA Briceglieri F.D., Eggermont E., Friedrich U., Haas O.A., Kobelt A.,
 RA Leroy J.G., van Maldergem L., Michel E., Mitulla B., Pfeiffer R.A.,
 RA Schinzel A., Schmidt H., Scherer G.;
 RT "Mutational analysis of the SOX9 gene in campomelic dysplasia and
 RT autosomal sex reversal: lack of genotype/phenotype correlations.";
 RL Hum. Mol. Genet. 6:91-98(1997).
 RN [6]
 RP VARIANT CD SER-112.
 RX MEDLINE=98112438; PubMed=9452059;
 RA Goji K., Nishijima E., Tsugawa C., Nishio H., Pokharel R.K.,
 RA Matsuo M.;
 RT "Novel missense mutation in the HMG box of SOX9 gene in a Japanese XY
 RT male resulted in campomelic dysplasia and severe defect in
 RT masculinization.";
 RL Hum. Mutat. Suppl. 1:S114-S116(1998).
 RN [7]
 RP VARIANTS CD L-112; V-119; Y-165 AND R-170, AND 3D-STRUCTURE MODELING.
 RX MEDLINE=99377034; PubMed=10446171;
 RA McDowall S., Argentaro A., Ranganathan S., Weller P., Mertin S.,
 RA Mansour S., Tolmie J., Harley V.;
 RT "Functional and structural studies of wild type SOX9 and mutations
 RT causing campomelic dysplasia.";
 RL J. Biol. Chem. 274:24023-24030(1999).
 RN [8]
 RP VARIANTS CD LEU-154 AND THR-158.
 RA Preiss A., Argentaro A., Barroso I., Schafer A.J., Clayton A.H.,
 RA Ogata T., Harley V.R.;
 RT "Functional analysis of two novel SOX9 mutations causing campomelic
 RT dysplasia";
 RL unpublished observations (JAN-2000).
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE NORMAL SKELETAL
 CC DEVELOPMENT. MAY REGULATE THE EXPRESSION OF OTHER GENES INVOLVED
 CC IN CHONDROGENESIS BY ACTING AS A TRANSCRIPTION FACTOR FOR THESE
 CC GENES.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- DISEASE: DEFECTS IN SOX9 ARE THE CAUSE OF CAMPOMELIC DYSPLASIA
 CC (CD). CD IS A RARE, OFTEN LETHAL, DOMINANTLY INHERITED, CONGENITAL
 CC OSTEOCHONDRODYSPLASIA, ASSOCIATED WITH MALE-TO-FEMALE AUTOSOMAL
 CC SEX REVERSAL IN TWO-THIRDS OF THE AFFECTED KARYOTYPIC MALES. A
 CC DISEASE OF THE NEWBORN CHARACTERIZED BY CONGENITAL BOWING AND
 CC ANGULATION OF LONG BONES, UNUSUALLY SMALL SCAPULAE, DEFORMED
 CC PELVIS AND SPINE AND A MISSING PAIR OF RIBS. CRANIOFACIAL DEFECTS
 CC SUCH AS CLEFT PALATE, MICROGNATHIA, FLAT FACE AND HYPERTELEORISM ARE
 CC COMMON. VARIOUS DEFECTS OF THE EAR ARE OFTEN EVIDENT, AFFECTING
 CC THE COCHLEA, MALLEUS INCUS, STAPES AND TYMPANUM. MOST PATIENTS DIE
 CC SOON AFTER BIRTH DUE TO RESPIRATORY DISTRESS WHICH HAS BEEN
 CC ATTRIBUTED TO HYPOPLASIA OF THE TRACHEOBRONCHIAL CARTILAGE AND
 CC SMALL THORACIC CAGE.
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 246629; CAA86598.1; -;
 CC EMBL: S74506; AAB32870.1; -;
 CC EMBL: S74504; AAB32870.1; JOINED.
 CC EMBL: S74505; AAB32870.1; JOINED.
 CC PDB: 1SX9; 15-JAN-00.
 CC PDB: 1S9M; 03-MAY-00.
 CC TRANSFAC: T01853; -;
 CC MIM: 114290; -;
 CC InterPro: IPR000910; HMG_12_box.

DR Pfam; PF00505; HMG_box; 1.
 KW SMART; SM00398; HMG; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein; 3D-structure;
 KW Disease mutation.
 FT DNA_BIND 105 173 HMG_BOX.
 FT DOMAIN 339 378 GLN/PRO-RICH.
 FT DOMAIN 342 346 POLY-PRO.
 FT VARIANT 108 108 P -> L (IN CD).
 FT VARIANT 112 112 /FTID=VAR_003735.
 FT VARIANT 112 112 F -> L (IN CD; LOSS OF DNA BINDING).
 FT VARIANT 112 112 /FTID=VAR_003736.
 FT VARIANT 119 119 F -> S (IN CD).
 FT VARIANT 119 119 /FTID=VAR_003737.
 FT VARIANT 119 119 A -> V (IN CD; ALMOST NO LOSS OF DNA
 FT BINDING).
 FT VARIANT 143 143 /FTID=VAR_003738.
 FT VARIANT 143 143 W -> R (IN CD).
 FT VARIANT 152 152 /FTID=VAR_003739.
 FT VARIANT 152 152 R -> P (IN CD).
 FT VARIANT 154 154 /FTID=VAR_003740.
 FT VARIANT 154 154 F -> L (IN CD; 19 FOLD REDUCTION IN DNA
 FT BINDING).
 FT VARIANT 158 158 /FTID=VAR_008529.
 FT VARIANT 158 158 A -> T (IN CD; 6 FOLD REDUCTION IN DNA
 FT BINDING).
 FT VARIANT 165 165 /FTID=VAR_008530.
 FT VARIANT 165 165 H -> Y (IN CD; LOSS OF DNA BINDING).
 FT VARIANT 170 170 /FTID=VAR_008531.
 FT VARIANT 170 170 P -> R (IN CD).
 FT VARIANT 354 356 /FTID=VAR_003741.
 FT VARIANT 354 356 MISSING (IN CD).
 FT SEQUENCE 509 AA; 56137 MW; 9289CFBB8D6631A2 CRC64;
 SQ

Query Match 5.1%; Score 93; DB 1; Length 509;
 Best Local Similarity 25.2%; Pred. No. 8.5;
 Matches 54; Conservative 27; Mismatches 67; Indels 66; Gaps 12;

QY 17 PPTDSSVTFIIILTMLFLGSLGAWGT--TSISTGSIIFSLKTLRSQHGGQVGLKVSPPRA 74
 Db 307 PATHQGVVTF-----GSYGISSTAAATPASAGHW-----MSKQA 341
 QY 75 QPLPAQPAPALAQYQSQPPQPTRWVAPRNNAFCQSGGA---GSDSNPGNVQPNNSA 131
 Db 342 PPPPPQPPQAPP---APQAPPQPPQ--AAPQPPAAPPQPPQAHTLTTLSEPGSQS--- 393
 QY 132 PSVESHPVLEKLAH-----SYNPKFEWNKSGRVFIKSYSEDDIHRKSIKYSTWCST 186
 Db 394 ---RTHIKTEQLSPSHVSEQQQHSPOQIAYS-----PFNLPHYSFS--YPPITRSQYDYT 443
 QY 187 EHGKNR-----LDSAFRCMS-SKGPVY 207
 Db 444 DHONSSSYSHAAGQGTGLYSTFTYWNPAQRPMT 477

RESULT 12
 SYN3_RAT
 ID SYN3_RAT STANDARD; PRT; 579 AA.
 AC O70441;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Synapsin III.
 GN SYN3
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98256247; PubMed=9593663;

FT DNA_BIND 58 86 MEF2-TYPE (POTENTIAL).
FT DOMAIN 254 257 POLY-PRO.
FT DOMAIN 288 293 POLY-GLU.
FT DOMAIN 419 423 POLY-HIS.
FT DOMAIN 448 455 POLY-SER.
SQ SEQUENCE 498 AA; 53724 MW; 590678D1BD1B3723 CRC64;

Query Match 5.1%; Score 92.5; DB 1; Length 498;
Best Local Similarity 25.7%; Pred. No. 9.1;
Matches 43; Conservative 19; Mismatches 50; Indels 55; Gaps 9;
Qy 15 TSPPTDSSVTETIILCT-----MLFLGSLGAW-----GTTSTIST-----GSIFS 53
Db 332 TAYNTDYSLTSAQLGFTSPGMLSLGQASAWQEHHLGQTTLSLVLVAGQLSQGSNLS 391
Qy 54 LKTLRSOHGGVGLKVSRRAPLPAOPPALAQ-----PQYQSPQOPQTRWVAPRNR-- 106
Db 392 INTNON-----INIK-SEPISPPRDRWTPSGFQHHHHHPQOQPPQPPQ-----PROEMG 442
Qy 107 -----NAAFQSGGAGSDSNS-----PGNVQPNAPSVE 135
Db 443 RSPVDSLSSSSSYDGDREDPRGDFHSPVILGRPANTEDRESFSVK 489

Search completed: July 24, 2002, 19:39:09
Job time: 233 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:34:56 ; Search time 29.22 Seconds
(without alignments)
2012.946 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MLRLGLLOGLTRFHTSPPTD.....CARNGRVETNNEGPVSYMF 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	1307.5	71.9	437	4	Q96S57	Q96S57	homo sapien
2	1307.5	71.9	466	4	Q96AN4	Q96AN4	homo sapien
3	1307.5	71.9	548	4	Q9BT17	Q9BT17	homo sapien
4	1307.5	71.9	559	4	Q9BYJ9	Q9BYJ9	homo sapien
5	997.5	54.9	570	4	Q9Y5A9	Q9Y5A9	homo sapien
6	978.5	53.8	579	4	Q9BUJ5	Q9BUJ5	homo sapien
7	978.5	53.8	579	11	Q91YT7	Q91YT7	mus musculus
8	785.5	43.2	543	4	Q96M25	Q96M25	homo sapien
9	673.5	37.0	700	5	Q9VBZ5	Q9VBZ5	homo sapien
10	600	33.0	197	4	Q9NX79	Q9NX79	homo sapien
11	546	30.0	503	10	Q9LDB8	Q9LDB8	arabidopsis
12	543	29.9	552	10	Q9LVM8	Q9LVM8	arabidopsis
13	521	28.7	493	10	Q9FNR2	Q9FNR2	arabidopsis
14	521	28.7	495	10	Q93ZP1	Q93ZP1	arabidopsis
15	521	28.7	528	10	Q9FPE7	Q9FPE7	arabidopsis
16	521	28.7	530	10	O64526	O64526	arabidopsis

17	511.5	28.1	667	10	Q9LJE5	Q9LJE5	arabidopsis
18	507	27.9	549	10	Q94A30	Q94A30	arabidopsis
19	507	27.9	580	10	Q9ZVU7	Q9ZVU7	arabidopsis
20	504	27.7	542	10	Q9C7F6	Q9C7F6	arabidopsis
21	502	27.6	1455	10	Q9LUT8	Q9LUT8	arabidopsis
22	485.5	26.7	664	10	Q9LNG4	Q9LNG4	arabidopsis
23	481.5	26.5	420	10	Q9SQR7	Q9SQR7	arabidopsis
24	465	25.6	708	10	Q94BR4	Q94BR4	oryza sativ
25	464	25.5	499	10	Q9ASM7	Q9ASM7	oryza sativ
26	441	24.3	425	10	O04503	O04503	arabidopsis
27	302.5	16.6	306	3	Q06390	Q06390	saccharomyc
28	233.5	12.8	359	10	Q9S262	Q9S262	arabidopsis
29	201.5	11.1	280	5	O77335	O77335	plasmodium
30	198.5	10.9	671	4	Q9H6S0	Q9H6S0	homo sapien
31	196.5	10.8	721	5	Q9Y172	Q9Y172	homo sapien
32	196.5	10.8	722	5	Q9VZ01	Q9VZ01	homo sapien
33	194	10.7	738	11	Q9QY02	Q9QY02	rat
34	188.5	10.4	658	4	Q96M07	Q96M07	homo sapien
35	186	10.2	712	11	O54729	O54729	rat
36	113	6.2	437	4	Q96QL2	Q96QL2	homo sapien
37	106.5	5.9	502	10	Q9ZT21	Q9ZT21	tritium ae
38	105	5.8	697	5	O61121	O61121	crithidia f
39	105	5.8	931	10	Q9FUY2	Q9FUY2	arabidopsis
40	103.5	5.7	1766	5	Q9V591	Q9V591	arabidopsis
41	102	5.6	457	5	Q95ZN5	Q95ZN5	caenorhabdi
42	102	5.6	471	2	Q9EZF8	Q9EZF8	streptococc
43	102	5.6	651	5	Q95ZNF	Q95ZNF	caenorhabdi
44	102	5.6	653	5	Q22553	Q22553	caenorhabdi
45	102	5.6	992	4	Q9C0L5	Q9C0L5	homo sapien

ALIGNMENTS

RESULT 1

ID Q96S57 PRELIMINARY; PRT; 437 AA.
AC Q96S57;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DERMATOMYOSITIS ASSOCIATED WITH CANCER PUTATIVE AUTOANTIGEN-1
DE (FRAGMENT).
GN DACA-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Onouchi H., Muro Y., Tomita Y.;
RT "Dermatomyositis Associated with Cancer Autoantigen.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055518; BAB62751.1; -;
FT NON_TER 1
SQ SEQUENCE 437 AA; 47281 MW; CA10E70E09BCE47 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 437;
Best Local Similarity 94.6%; Pred. No. 4.5e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

Qy	74	AQPLPAQPALAQPQYQSQPPQPPQTRWVAPRNRNNAAFQSGGAGSDSPGNVQPSNAPS	133
Db	180	AQPLPAQPALAQPQYQSQPPQPPQTRWVAPRNRNNAAFQSGGAGSDSPGNVQPSNAPS	239
Qy	134	VESHPVLEKLAHSAHYNKPEFENLKSGRVFIKSYSEDDIHRISKYISWSTEGHKNRL	193
Db	240	VESHPVLEKLAHSAHYNKPEFENLKSGRVFIKSYSEDDIHRISKYISWSTEGHKNRL	299
Qy	194	DSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGACVWSQDKWKFQVWIFV	253
Db	300	DSAFRCMSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGACVWSQDKWKFQVWIFV	359

QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003681; AAH03681.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 466;
Best Local Similarity 94.6%; Pred. No. 4.9e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
AC Q96AN4; PRELIMINARY; PRT; 466 AA.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 50.6 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016920; AAH16920.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50594 MW; 7484D2CD1BCE7D22 CRC64;

Query Match 71.9%; Score 1307.5; DB 4; Length 466;
Best Local Similarity 94.6%; Pred. No. 4.9e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
DB 209 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 268
QY 134 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
DB 269 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 328
QY 194 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
DB 329 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 388
QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
DB 389 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 448
QY 314 RQRRRCWCAARNGRVETNNE 332
DB 449 RQEEEE-VVRKERQSRNQ 466

Query Match 71.9%; Score 1307.5; DB 4; Length 548;
Best Local Similarity 94.6%; Pred. No. 6.1e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
DB 302 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 361
QY 134 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
DB 362 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 421
QY 194 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
DB 422 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 481
QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313

Query Match 71.9%; Score 1307.5; DB 4; Length 548;
Best Local Similarity 94.6%; Pred. No. 6.1e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
DB 291 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 350
QY 134 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
DB 351 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 410
QY 194 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
DB 411 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 470
QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313
DB 471 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 530
QY 314 RQRRRCWCAARNGRVETNNE 332
DB 531 RQEEEE-VVRKERQSRNQ 548

Query Match 71.9%; Score 1307.5; DB 4; Length 559;
Best Local Similarity 94.6%; Pred. No. 6.3e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
DB 302 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 361
QY 134 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
DB 362 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 421
QY 194 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
DB 422 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 481
QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313

Query Match 71.9%; Score 1307.5; DB 4; Length 559;
Best Local Similarity 94.6%; Pred. No. 6.3e-109;
Matches 245; Conservative 1; Mismatches 12; Indels 1; Gaps 1;

QY 74 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 133
DB 302 AQLPAPALAPQYQSQPPQPTRWVAPRNRNAAFGSGGAGSDNSPGNVQPNAPS 361
QY 134 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 193
DB 362 VESHPVLEKLAHAHSYNPKFEFNLKSGRVFIISKYSEDDIHRSIKYSIWCSTEHGNKRL 421
QY 194 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 253
DB 422 DSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVMSQDKWKGFQVQWIFV 481
QY 254 KDVPNNQLRHRLNNDNKPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 313


```
|||||
Db 482 KDVPNNQLRHRLNNDKPVTSRDQVPLEKAKQVLKIISSYKHTTSIFDDFAHYEK 541
QY 314 RRRRRRCARNGRVETNNE 332
Db 542 ROEEEE-VVRRKQSRNKQ 559

RESULT 5
Q9Y5A9 ID Q9Y5A9 PRELIMINARY; PRT; 570 AA.
AC Q9Y5A9
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NY-REN-2 ANTIGEN.
GN HGRG8.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA Old L.J.;
RT *Antigens recognized by autologous antibody in patients with renal-
RT cell carcinoma*;
RL Int. J. Cancer 0:0-0(1999).
RN 12
RP SEQUENCE FROM N.A.
RA Roberts T.P., Wright A., Wahab N.A., Weston B.S., Mason R.M.;
RA *Gene which is selectively expressed in hyperglycaemia.*;
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF155095; ARD42861.1; -
DR EMBL; AF192968; AAF08813.1; -
SQ SEQUENCE 570 AA; 61320 MW; 91FCFA7E508869E4 CRC64;

Query Match 54.9%; Score 997.5; DB 4; Length 570;
Best Local Similarity 75.7%; Pred. No. 4.5e-81;
Matches 190; Conservative 19; Mismatches 35; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQPAQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCSTEGHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYNIMWCSTEGHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYCTSAQVWSQDKWKGFQVOW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGRFDVRW 499
QY 251 IFVKDVPNNQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
Db 500 IFVKDVPNSQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIIASYKHTTSIFDDFESH 559

RESULT 6
Q9BUJ5 ID Q9BUJ5 PRELIMINARY; PRT; 579 AA.
AC Q9BUJ5
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HIGH-GLUCOSE-REGULATED PROTEIN 8.
OS Homo sapiens (human).

Query Match 54.9%; Score 997.5; DB 4; Length 570;
Best Local Similarity 75.7%; Pred. No. 4.5e-81;
Matches 190; Conservative 19; Mismatches 35; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQPAQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCSTEGHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYNIMWCSTEGHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYCTSAQVWSQDKWKGFQVOW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGRFDVRW 499
QY 251 IFVKDVPNNQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
Db 500 IFVKDVPNSQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIIASYKHTTSIFDDFESH 559

RESULT 7
Q91YT7 ID Q91YT7 PRELIMINARY; PRT; 579 AA.
AC Q91YT7
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE HYPOTHETICAL 62.3 KDA PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014797; AAH14797.1; -
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 62280 MW; 9D79DF630593C7AB CRC64;

Query Match 53.8%; Score 978.5; DB 11; Length 579;
Best Local Similarity 75.9%; Pred. No. 2.3e-79;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQPAQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCSTEGHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYNIMWCSTEGHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYCTSAQVWSQDKWKGFQVOW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGRFDVRW 499
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA, CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002559; AAH02559.1; -
SQ SEQUENCE 579 AA; 62334 MW; BF3959B5561A464E CRC64;
```

```
Query Match 53.8%; Score 978.5; DB 4; Length 579;
Best Local Similarity 75.9%; Pred. No. 2.3e-79;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQPAQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCSTEGHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYNIMWCSTEGHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYCTSAQVWSQDKWKGFQVOW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGRFDVRW 499
QY 251 IFVKDVPNNQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
Db 500 IFVKDVPNSQLRHRLNNDKPVTSNRDQVPLEKAKQVLKIIASYKHTTSIFDDFESH 559
QY 311 YEKRO 315
Db 560 YEKRO 564
```

RESULT 7

```
Q91YT7 ID Q91YT7 PRELIMINARY; PRT; 579 AA.
AC Q91YT7
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE HYPOTHETICAL 62.3 KDA PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014797; AAH14797.1; -
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 62280 MW; 9D79DF630593C7AB CRC64;
```

Query Match

```
Best Local Similarity 75.9%; Pred. No. 2.3e-79;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;

QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGGAGSDNSPGNVQ--PNS 130
Db 325 QTQPLPPPPPPQPAQLSVQOQAAQP--TRWVAPRNRGSGFGHN--GVDGNGVGQSQAGSGS 380
QY 131 APSVESHVPLEKKAHNSYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCSTEGHN 190
Db 381 TPS-EPHPVLEKLSINNYNPKDFDNLKGRVFIKSYSEDDIHRSIKYNIMWCSTEGHN 439
QY 191 KRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYCTSAQVWSQDKWKGFQVOW 250
Db 440 KRLDAAYRSMNGKGPVYLLFSVNGSGHFCGVAEMKSAVDYNTCAGVWSQDKWKGRFDVRW 499
```

```
QY 251 IFVKDVPNNQLRHRLNNDNKPVTNSRDQTEVPLEKAKQVLKIISYKHTTSIFDDFAH 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 500 IFVKDVPNSQLRHRLNNDNKPVTNSRDQTEVPLEKAKQVLKIISYKHTTSIFDDFSH 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 311 YEKRO 315
|||||
Db 560 YEKRO 564
|||||

RESULT 8
Q96M25 PRELIMINARY: PRT: 543 AA.
AC Q96M25;
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE CDNA FLJ131657 FIS, CLONE NT2R12004304, MODERATELY SIMILAR TO HOMO
DE SAPIENS NT-REN-2 ANTIGEN MRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK056219; BAB71122.1; -- 6D3BCE5B83F9BAF5 CRC64;
SQ SEQUENCE 543 AA; 59032 MW; 6D3BCE5B83F9BAF5 CRC64;

Query Match 43.2%; Score 785.5; DB 4; Length 543;
Best Local Similarity 59.9%; Pred. No. 4.5e-62;
Matches 154; Conservative 27; Mismatches 51; Indels 25; Gaps 4;

QY 37 SLGANGTTSISTGTSIFSLKTLRSQGGQVGLKYSRRAQP-----LPAQPPALAAQPOY 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 NIGTD-----ERGSYVAKPPTQVLPPTIIQQPQPLIQPPLVQSOLPQQPQQPQQ 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 90 QSPQOPP-----QTRWVAPRNRAAFGSGGAGSDSNSGNVQPNAP-SVE 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 QGPGPQAQPHQVQPOQOOLQNRWVAPRNRGAGFNQNGAGSENGLGVVVPVSASPSSVE 390
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 136 SHPVLEKLAHNSYNPKFEFWNLKSGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDS 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 391 VHPVLEKLAHNSYNPKFDNLKNGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDA 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 196 APRCMSSGPGVLLFSVNGSHFCGVAEMKSPVDYGTSGAGVWSQDKWKGKFDVQVIFVKD 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 451 AYVRLNGKPLVLLFSVNGSHFCGVAEMKSPVDYGTSGAGVWSQDKWKGKFDVQVIFVKD 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 256 VPNNQLRHRLNNDNKP 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 VPNNQLRHRLNNDNKP 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
Q9VBZ5 PRELIMINARY: PRT: 700 AA.
AC Q9VBZ5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CG6422 PROTEIN.
GN CG6422.
```

```
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfiffner B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003750; AAF36381.1; --.
DR FlyBase: FBgn0039261; CG6422.
SQ SEQUENCE 700 AA; 78916 MW; FD1D6C53F769C363 CRC64;

Query Match 37.0%; Score 673.5; DB 5; Length 700;
Best Local Similarity 48.7%; Pred. No. 7.3e-52;
Matches 150; Conservative 35; Mismatches 84; Indels 39; Gaps 9;

QY 70 SRPRAQPL-----PAQPPALAAQPO-----YQSPQPPQTRW-----VAPRNRAAFGOS 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 260 SRKYSPTGPVRNLGPPQPVHHAAPRATNAGPEGPPNARRHDGPHPSRSERSGNYS 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 114 GGAGS-----DSNSPGNVQPNAPS---VESHPVLEKLAHNSYNPKFEFWNLK 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 320 SFRGEFESAKFEYRDENQSRPEATSATELPVDSQLVLDLKDKNYNPKVLDD--LK 377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 160 ---SGRVFIKSYSEDDIHRSIKYSIMCSTEHGKRLDSAPRCMSK-K-CPVYLLFSVNGS 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 KAGSARFFVIKSYSEDDIHRSIKYSIMCSTDHGNKRLDDAFKRIHEEGGNIMLFFSVNGS 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 216 GHFCGVAEMKSPVDYGTSGAGVWSQDKWKGKFDVQVIFVKDVPNNQLRHRLNNDNKPVT 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 438 GHFCGVAEMKSPVDYGTSGAGVWSQDKWKGKFDVQVIFVKDVPNNQLRHRLNNDNKPVT 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 276 NSRDQTEVPLEKAKQVLKIISYKHTTSIFDDFAIYKQRRRRWCAR---NGRVETNNE 332
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Matches	104;	Conservative	43;	Mismatches	58;	Indels	22;	Gaps	4
Qy	95	PPOTRWVAPNRNAAFGGSGAGSDNSPGNVQNSAPSVESHVPVLEKLAHSYNPKF	154						
Db	302	PRASSRVKSNSSKPCSTIGDSASDSSTAG---PN-PSLYNHP-----EF	342						
Qy	155	EWNLKSGRVFIKSYSEDDIHRSIKYSIWCSTEHGKRLDSAFR---CMSSKGPVYLLFS	211						
Db	343	VTDYKNAKFFVKSFSEDNVHRSIKYNVWASTPHGNKKLDTAYDAEAKMGCKPIFLFS	402						
Qy	212	VNGSGHFCGVAEMKSPVDYGTSGAGVWQDKWKKFQVQWIFVKDVPNNQLRHIRELNDN	271						
Db	403	VNASGQFCGVSEMGVDPFEKDAQWQDWSGQFPVKWHIVKIDIPNRRFCHILLQNNND	462						
Qy	272	KPVTNSRDTQVPLEKAKOVLLKISSYKHTTSIFDDFAHYEKRRRR	318						
Db	463	KPVTHSRDSQEVKURQGIEMLRIFKEYEHTSIILDDFGYDELEGQK	509						
RESULT	13								
Q9FN92	ID	Q9FN92	PRELIMINARY;	PRT;	493	AA.			
AC	Q9FN92;								
DT	01-MAR-2001	(TREMBLrel. 16, Created)							
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)							
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)							
OS	SIMILARITY TO UNKNOWN PROTEIN								
OS	Arabidopsis thaliana (Mouse-ear cress).								
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;								
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsis.								
NCBI	TaxID=3702;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=COLUMBIA;								
RX	MEDLINE=98069011; PubMed=9405937;								
RA	Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,								
RA	Tabata S.;								
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. II.								
RT	Sequence features of the regions of 1,044,062 bp covered by thirteen								
RL	physically assigned Pl clones.";								
RL	DNA Res. 4:291-300(1997).								
DR	EMBL; AB006696; BAB10365.1; .								
SQ	SEQUENCE 493 AA; 54984 MW; E5539254252DEAF4 CRC64;								
Query Match	28.7%; Score 521; DB 10; Length 493;								
Best Local Similarity	44.9%; Pred. No. 2.2e-38;								
Matches	114;	Conservative	32;	Mismatches	80;	Indels	28;	Gaps	6;
Qy	89	YQSPQPPQTFRWA-----PRNRNAAFQSGGAGSDNSPG-----NVQPNs----	133						
Db	166	YDSWKYMP--NWYAVNNTYKPRNGYHGKNIENEMNMRGPRAKFNSQDGSKVMAVS	223						
Qy	134	VESHVPVLEKLAH-----SYNPEFEWNLKSGRVEIKSYSEDDIHRSIKYSIWCs	185						
Db	224	LKEORVTEKLSDEVSLLDPKDNKIDFPPTYTEAFVTKSYSEDDIHRSIKYSIWSs	283						
Qy	186	TEHGNKRLDSAFRCMSKSG---PYVLLFSVNGSGHFCGVAEMKSPVDYGTSGAGVWQDKW	242						
Db	284	TPNGKKLDASYNEAKQKSDGCPVFLFSVNTSGQFVGLAEWVGPDVFNKTVETWQDQKW	343						
Qy	243	KGKEDVQWIFVKVQVNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKOVLLKISSYKHTT	302						
Db	344	IGCPFVKWHFVKDIPNSSLRHITLLENNENKPVNTSRDTQEVKLEGGIKVIFKDHASKT	403						
Qy	303	SIFDDFAHYEKRR	316						
Db	404	CILDFFEYENRQK	417						
RESULT	14								
Q93ZP1	ID	Q93ZP1	PRELIMINARY;	PRT;	493	AA.			

ID	Q93ZP1	PRELIMINARY;	PRT;	495 AA.
AC	Q93ZP1;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	AF5G61020/MAF19_20.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi			
OX	NCBI_TaxID=3702;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RA	Chen R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,			
RA	Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,			
RA	Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,			
RA	Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.			
RA	Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.			
RA	Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,			
RA	Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,			
RA	Ecker J.R.;			
RT	"Arabidopsis cDNA clones.";			
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY056421; AAL08277.1; -			
SQ	SEQUENCE 495 AA; 55211 MW; 55B3EF0C6B6F5CC8 CRC64;			
Query Match	28.7%;	Score	521;	DB 10; Length 495;
Best Local Similarity	44.9%;	Pred. No.	2.3e-38;	
Matches	114;	Conservative	32;	Mismatches 80; Indels 28; Gaps
Qy	89 YQSPQOPQTRFWA-----PRNRNAAFGSGGAGSDNSPG-----NVQPSN---APS 133			
Db	168 YDSWKYMP--NWAVNNTYKPRNGYHGYGKENTEGLENNMRGPRAK3FNSQDGSKYMAVS 225			
Qy	134 VESHPVLEKLAAH-----SYNPKPEFENLKSGRVFIKSYSEDDIHRSIKYSIWCS 185			
Db	226 LKEQVTEKLSDEVSLLDPKDYNKIDFPETYTEAKFYIKSYSEDDIHKSIKYSWSS 285			
Qy	186 TEHGNKRLDSAFRCWMSKG---PVYLLFSVNGSHGFCGVAEMKSPVDYGTGSAGVWSQDKW 242			
Db	286 TPNGNKKLDASYNKAKGDCGCPVLLFSYNTSCQFVGLAEMVGPVDFNKTVEYWOODKW 345			
Qy	243 KGFQDVQWIFVKVPNQLRHIRELNNKPVNRSRDTQEVPLEKAKOVLKLIISYKHIT 302			
Db	346 IGCFFPKWHFVKDIPNSSLRHITLNNENKPVNRSRDTQEVKLEQGIKVIKIPKDHASKT 405			
Qy	303 SIFDDFAHYEKRRQ 316			
Db	406 CILDDFFEYENRQK 419			
RESULT 15				
Q9FPE7				
ID	Q9FPE7	PRELIMINARY;	PRT;	528 AA.
AC	Q9FPE7;			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL 59.5 KDA PROTEIN.			
GN	YUP8H12R.13.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,			
RA	Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,			
RA	Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,			
RA	Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,			
RA	Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,			

Job time: 231 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:32:51 ; Search time 17.76 Seconds
(without alignments)
1839.548 Million cell updates/sec

Title: US-09-877-633-1

Perfect score: 1818

Sequence: 1 MRLRGLLQGLTFRHFTSPPTD.....CARNGRVETNNEGEPSVMF 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 28138_seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	521	28.7	530	2 T01030	hypothetical prote
2	507	27.9	580	2 C96597	Rubisco subunit bi
3	504	27.7	542	2 A86405	unknown protein [i
4	441	24.3	425	2 C86232	hypothetical prote
5	302.5	16.6	306	2 S61169	hypothetical prote
6	233.5	12.8	359	2 T06604	hypothetical prote
7	201.5	11.1	280	2 T18443	hypothetical prote
8	107.5	5.9	1533	2 A46221	abdominal segment
9	102	5.6	653	2 T34356	hypothetical prote
10	97.5	5.4	292	2 T18584	F15C11.2 - Caenorh
11	97.5	5.4	502	2 T18562	hypothetical prote
12	97.5	5.4	508	2 S54264	glycoprotein gc -
13	97	5.3	980	1 S45444	BEM1 protein-bindi
14	96	5.3	367	1 OZZOZY	circumsporozoite p
15	96	5.3	892	2 T09193	ataxin 7 - human
16	95	5.2	1168	1 MMAXIC	myosin heavy chain
17	94.5	5.2	602	2 S60052	calcium-dependent
18	94	5.2	520	2 G86414	probable protein k
19	94	5.2	930	2 T08588	hypothetical prote
20	93.5	5.1	1074	2 T24877	hypothetical prote
21	93.5	5.1	1076	2 T24887	hypothetical prote
22	93	5.1	507	2 S25831	myocyte-specific e
23	93	5.1	509	2 A55204	transcription fact
24	93	5.1	1329	2 A64828	cell division prot
25	93	5.1	2215	2 T16871	hypothetical prote
26	92.5	5.1	667	2 T09482	ring finger protel
27	92	5.1	283	2 T08735	hypothetical prote
28	92	5.1	491	2 H84477	probable PttA-like
29	92	5.1	491	2 T12501	hypothetical prote

30 91.5 5.0 645 2 A71416
31 91.5 5.0 785 2 S54016
32 91 5.0 1342 2 G90750
33 91 5.0 1342 2 E85614
34 91 5.0 1353 2 T00347
35 90.5 5.0 741 2 D81798
36 90.5 5.0 877 1 A25962
37 90.5 5.0 1047 2 A59246
38 90.5 5.0 1181 2 C86349
39 90.5 5.0 2515 2 A41519
40 90 5.0 1053 2 T07965
41 89.5 4.9 502 2 T08776
42 89.5 4.9 856 2 T13159
43 89.5 4.9 1460 2 D81675
44 89 4.9 457 2 T04226
45 89 4.9 849 2 S61962

ALIGNMENTS

RESULT 1

T01030

hypothetical protein YUP8H12R.13 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999

C:Accession: T01030

R:Rheologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K

Oefner, P.; Davis, R.W.

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A:Reference number: Z14227

A:Accession: T01030

A:Status: translated from CB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-530 <RHE>

A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152559; GSPDB:GN000059; ATSP:YU

C:Genetics:

A:Gene: ATSP:YUP8H12R.13

A:Map position: 1

A:Introns: 377/3; 77/1; 100/3; 219/3; 380/3; 451/3; 499/3

Query Match 28.7%; Score 521; DB 2; Length 530;

Best Local Similarity 43.6%; Pred. No. 1.3e-32;

Matches 125; Conservative 28; Mismatches 94; Indels 40; Gaps 7;

QY 61 HGGQVGLKVRPRAQPLPAQPPALAAQP--QYQSPQQP-----PQTRWVAPRNR-- 106

DB 208 HGMGDRPKTPRKASQNSYAPPPLLNQEKGRITAYPMDPVKKSGALNRDETEKAKARTKEN 267

QY 107 -----NAAFGQ-----SGAGSADS--NSPGNVQPNAPSVPESHVPLEKLAHSHYKPKFEW 156

DB 268 GTSMNDLANGQDHITNGECESCSLDAEGNERSNGVSVIRR-----DOYNLPSFQT 318

QY 157 NLKSGRVFLKSYSEDDIHSIKYSIWCSTHGKRLDLSAPR-----CMSSGQPVVLL 209

DB 319 KYEARIFVTKSYSEDDIHSIKYVWSSTLNGNKKLDSAYOESQKRAADSKGCPFLP 378

QY 210 FSVNGSGHFCGVAEMKSPVDYGTSGAVWSQDKWKGFQVQWIFVKDVPNNLRHRLENN 269

DB 379 FSVNASQFCGVAEMIGRDVYEKSMFQQDKWTGYFPVKWHIHKDVPNQLRHILENN 438

QY 270 DNKPVVNSRDTQEVPLEKAKQVLKLIISYKHTTTSIFDDFAHYKQR 316

DB 439 ENKPVVNSRDTQEVRLPQGNVLELIFKNYAARTSILDDDFDYENREK 485

RESULT 2

C96597

Rubisco subunit binding-protein beta subunit [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

```

C:Genetics:
A;Map position: 1

Query Match 27.7%; Score 504; DB 2; Length 542;
Best Local Similarity 43.9%; Pred. No. 2.7e-31;
Matches 108; Conservative 34; Mismatches 68; Indels 38; Gaps:

Qy 106 RNAAFGSGGA---GSDSNSPG-----NVQPNAPSYESH-----PVLEKLLKAAHSYNPK 152
Db 243 REQAYACKRPSPASSSSTWENDYLNPLDARSESYNDSHCAMLDMLTESNRGPR 302

Qy 153 EFEWNLKS-----GRVFIISKSEDDIHRSIKYSIWCSTEGHN 190
Db 303 ASRLNSKSMISYDRVDRFCQELLQSFRDAKFEVIKSYSEDNVHKSIKHCVMASTKNGN 362

Qy 191 KRLDSAPRCMSSKG---PVYLLSVANGSGHFCGVAEMKSPVDYGTSGAGVWSODKHWGKGP 247
Db 363 KKLDAAYREAKKDVACPVFLFSVNASOFCGVAEMGVDFNTSVYEQDQRWSGHP 422

Qy 248 VOWLFYKDVNNOLRHRLLENNDKNPVTNSRDTQEVPLEKAKOVLIKTISSYKHTTSIFDD 307
Db 423 VOWLIVKDVNSLFRHIIIESNDKNPVTNSRDTQEVGLEKEMLDIFISCERSSILDD 482

Qy 308 FAHYEKRQ 315
Db 483 FNFYEERQ 490

RESULT 4
C86232
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86232
R:Theologits, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.L.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykhi,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Mafti,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86232
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <STO>
A:Cross-references: GB:AE005172; NID:g2160172; PIDN:AAB60735.1; GSPDB:GN00
C:Genetics:
A;Map position: 1

Query Match 24.3%; Score 441; DB 2; Length 425;
Best Local Similarity 41.2%; Pred. No. 1.5e-26;
Matches 89; Conservative 35; Mismatches 64; Indels 28; Gaps:

Qy 132 PSVESHVPVLEKLKAA-----HSYNPKFEFWNLKSGRVFIKSYSEDDIHRSIKYSIWCST 186
Db 136 PPSSESLPKNNFSFALALRREYNLPDQTDYEDAKFFVIKSYSEDDVHKSIKYSVWSST 195

Qy 187 EHGKRLDLSAFRCMSS-----KGPVYLLFS-----VNGSGHFCGVAE 223
Db 196 INGNKLLDAAFRDAETKTLDEGDKRKRIPTFLPSGVLLVSLMISYPSFIQYNASHQFVGLAE 255

Qy 224 MKSPVDYGTSGAGVWSODKWKGGEDVQWIFVKDVPNNOLRHRLLENNDKNPVTNSRDTQEV 283
Db 256 MGVYDFENKOLDPQVDKWSGFTFVEMHVVKDIPNNELRHILLDNEDKPVVTHRTROI 315

Qy 284 PLEKAKOVLIKTISSYKHTTSIFDDFAHYEKRRRR 319

```


A:Molecule type: DNA
A:Residues: 1-502 <WIL>
A:Cross-references: EMBL:Z98262; PIDN:CAB10932.1; GSPDB:GN00019; CESP:F15C11
A:Experimental source: clone VF15C11
R:Wilkinson, J.
Submitted to the EMBL Data Library, April 1996
A:Reference number: Z19352
A:Accession: T20970
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-502 <WIL2>
A:Cross-references: EMBL:Z71260; PIDN:CAA95799.1; GSPDB:GN00019; CESP:F15C11
C:Genetics:
A:Experimental source: clone F15C11
A:Gene: CESP:F15C11.2
A:Map position: 1

A; Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match 5.4%; Score 97.5; DB 2; Length 502;
Best Local Similarity 26.9%; Pred. No. 7.8; Indels 23; Gaps 7;
Matches 45; Conservative 25; Mismatches 74; Indels 23; Gaps 7;

Qy 11 LRFHTSPPT-----DSSVTEIILCTMLFLGSLGAWGTTSTSGTSPSLKTLRSQH 61
Db 8 IKVHKSPSNKYDVEIADASVSE-LKDKVLVFPVTANKEQVCIIYTGKILKDEETLQH 66
Qy 62 ---GGVGLKYSRPAQLPAQPPALAAQOYQSPQPPQTRWVAPRNRNNAAFQSGGAG 117
Db 67 KIADGHTVHL-VIRNQAARTPA--PAAATPTASSAPSSNPTPPSOPNPTNNPFAAMGGM 123
Qy 118 SDSNPGNVOPNSAPSVESHVPLEKLAHAHSYNPKFEWNLKSGRVF 164
Db 124 SPADILNN--PDAMRSVNDNPTIQQLLG-----NPEFWRTIIQSNPQF 164

RESULT 12
S54264
glycoprotein gC - bovine herpesvirus 1
C; Species: bovine herpesvirus 1
C; Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
R; Hecht, P.; Engels, M.; Loeffe, E.; Ackermann, M.
submitted to the EMBL Data Library, May 1995
A; Description: Comparison of the glycoprotein gC genes of bovine and caprine herpesvirus
A; Reference number: S54264
A; Accession: S54264
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-508 <HPC>
A; Cross-references: EMBL:Z49223; NID:9804966; PIDN:CAA89198.1; PID:9804967
C; Superfamily: herpesvirus glycoprotein F
C; Keywords: glycoprotein

Query Match 5.4%; Score 97.5; DB 2; Length 508;
Best Local Similarity 24.9%; Pred. No. 7.9; Indels 81; Gaps 11;
Matches 57; Conservative 15; Mismatches 76; Indels 81; Gaps 11;

Qy 16 SPPTDSSVTEIILCTMLFLGSLGAWGTTSTSGTSPSLKTLRSQHGQVGLKYSRPAQ 75
Db 34 SPPSPSPPTET-----ESSAGTTGATPTTPNS-----PDAT 64
Qy 76 PLPAQPPALA-----QPQYQSPQPPQTRWVAPRNRNNAAFQSGGAGSDNS----- 122
Db 65 PEDSTPGATTGGTPEPPSVSEHDPVTN-STPPPPAPPEDGRPGGAGNASRDGRPSGGGR 123
Qy 123 PGNVQPNAPSVPESHVPLEKLAHAHSYNPKFEWNLKSGRVFIITKSYSE-----DD 173
Db 124 PRPQPSKAP-----PKERKMWL-CEREAVAASAEPLVYHCGVADN 164
Qy 174 IHRSIKYSIMCS-----TEHGKRLDSAFRCMSKGPVYLLFSV-NGS 215
Db 165 ATGGARLELWQVRGRFSTRGD---DEAVRNPPFPAPPVLLFVVQNGS 210

RESULT 13
S45444
BEM1 protein-binding protein BOB1 - yeast (Saccharomyces cerevisiae)
N; Alternate names: protein YBL0717; protein YBL085W
C; Species: Saccharomyces cerevisiae
C; Date: 09-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
R; Bender, A.; Bender, L.; Kokojan, V.
submitted to the EMBL Data Library, April 1994
A; Description: Yeast Bob1p (Bem1p-binding protein) binds to the SH3 domain-containing pr
A; Reference number: S45444
A; Accession: S45444
A; Molecule type: DNA

A; Residues: 1-980 <BEN>
A; Cross-references: EMBL:L31406; NID:9829041; PIDN:AAB08439.1; PID:9466436
R; Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
submitted to the EMBL Data Library, May 1994
A; Description: Sequence analysis of a 78.6 kb segment of the left end of Saccharomyces
A; Reference number: S45387
A; Accession: S45421
A; Molecule type: DNA
A; Residues: 1-980 <ORE>
A; Cross-references: EMBL:X79489; NID:9496661; PIDN:CAA56021.1; PID:9496694
R; Domdey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.
submitted to the Protein Sequence Database, August 1994
A; Reference number: S45816
A; Accession: S45826
A; Molecule type: DNA
A; Residues: 1-980 <DOM>
A; Cross-references: EMBL:Z35846; NID:9536137; PIDN:CAA84906.1; PID:9536138; GSPDB:GNO
R; Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
Yeast 11, 1103-1112, 1995
A; Title: Sequence analysis of a 78.6 kb segment of the left end of Saccharomyces cere
A; Reference number: S59184; MUID:96076635
A; Accession: S59218
A; Status: nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-980 <OBW>
A; Cross-references: EMBL:X79489; NID:9496661; PIDN:CAA56021.1; PID:9496694
A; Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
C; Genetics:
A; Gene: SGD:BOI1; BOB1; MIPS:YBL085W
A; Cross-references: MIPS:YBL085W; SGD:S0000181
A; Map position: 2L
C; Superfamily: BEM1 protein-binding protein BOB1; pleckstrin repeat homology; SAM hom
F:20-72/Domain: SH3 homology <SH3>
F:225-291/Domain: SAM homology <SAM>

Query Match 5.3%; Score 97; DB 1; Length 980;
Best Local Similarity 19.9%; Pred. No. 21;
Matches 61; Conservative 45; Mismatches 112; Indels 88; Gaps 14;

Qy 50 SIFSLKTLRSOHGQVGLKYSRPAQPLPAQPPALAAQOYQSP-----QQPQT 98
Db 383 NLFADKQIFESPG-----RAPKPPSPSPVQPP-----QSPSFNNRYTNNARFPQT 430
Qy 99 RWAPRNRNNAAFQSGGAGSDNSPGNVOPNSAPSVESHVPLEKLAHAHSYNPKFEWNL 158
Db 431 TY-PPKNNKPTVYNG-----LIPNSSTSDNSTGKFKPPAMNGHDSNRKTTL 478
Qy 159 KSGRVFIKYSSEDI-----HRSIKYSIMCSSTEHGKRLDSAFRCMS 201
Db 479 TSATIPSTINTVNTDESPLAISNISSNATSHHPNRSVYNNHKKRTESGSSFVD-LFNRI 537
Qy 202 SKGPVYLLFSYNGSGHFCGVAEMKSPVDYGTSGAGVWSODKWKGFVDQWIFVKDVPNNQL 261
Db 538 MLSPVKSFDEE-----ETKQPSK--ASRAVFDASARKKSSYG-----HSRDASLSEM 582
Qy 262 RHIRLENNNDKPVTN--SRDQEVPLEKAKOVLT-----ISSYKHTTTSIEDD 307
Db 583 K-----KHRRNSSILSFFSSKSSQSNPTSPTKQTFTIDPAKMTSHRSRQSNYSYHARS--QS 636
Qy 308 FAHYEK 313
Db 637 YSHSRK 642

RESULT 14
OZZQMY
circumsporozoite protein precursor - Plasmodium yoelii
N; Alternate names: sporozoite surface antigen
C; Species: Plasmodium yoelii
C; Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C; Accession: A26271
R; Lal, A.A.; de la Cruz, V.F.; Welsh, J.A.; Charoenvit, Y.; Maloy, W.L.; McCutchan, T

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:32:01 ; Search time 13.18 Seconds
(without alignments)
630.099 Million cell updates/sec

Title: US-09-877-633-1
Perfect score: 1818
Sequence: 1 MRLRGLLQGLRFTSPPTD.....CARNGRVETNNEGEFVSVMF 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1818	100.0	340	2	US-08-933-750C-26
2	1818	100.0	340	4	US-09-234-613-26
3	99	5.4	559	4	US-09-269-731-2
4	96	5.3	478	3	US-08-155-888-2
5	93	5.1	507	5	PCT-US93-08386-8
6	93	5.1	509	4	US-08-860-635A-21
7	93	5.1	509	4	US-09-281-476-21
8	90	5.0	363	3	US-08-586-165-7
9	88	4.8	345	3	US-08-718-738-2
10	88	4.8	345	4	US-09-221-844-2
11	87.5	4.8	580	3	US-08-906-865-1
12	86.5	4.8	521	2	US-08-682-847-4
13	85.5	4.7	188	1	US-08-842-255-131
14	85.5	4.7	198	1	US-08-397-633A-52
15	85.5	4.7	345	5	PCT-US93-0323A-2
16	85	4.7	355	4	US-08-630-915A-192
17	85	4.7	574	4	US-09-276-400-6
18	85	4.7	574	4	US-09-448-076-6
19	85	4.7	623	4	US-09-347-801-4
20	85	4.7	810	6	5200340-8
21	85	4.7	1494	3	US-08-755-587-186
22	84	4.6	339	1	US-08-248-629A-3
23	84	4.6	339	1	US-08-451-932-3
24	84	4.6	339	1	US-08-452-260-3
25	84	4.6	339	1	US-08-326-785-3
26	84	4.6	339	2	US-08-612-788-3
27	84	4.6	339	2	US-08-605-598B-3

28	84	4.6	339	2	US-08-429-743-3	Sequence 3, Appli
29	84	4.6	339	2	US-08-866-735-3	Sequence 3, Appli
30	84	4.6	339	3	US-09-066-028-3	Sequence 3, Appli
31	84	4.6	339	5	PCT-US95-05107-3	Sequence 3, Appli
32	84	4.6	352	2	US-08-612-788-40	Sequence 40, Appl
33	84	4.6	378	2	US-09-066-028-40	Sequence 40, Appl
34	84	4.6	378	2	US-08-612-788-42	Sequence 42, Appl
35	84	4.6	378	3	US-09-066-028-42	Sequence 42, Appl
36	84	4.6	378	4	US-09-206-059-1	Sequence 1, Appli
37	84	4.6	790	1	US-08-469-486-54	Sequence 54, Appl
38	84	4.6	790	2	US-08-469-658-54	Sequence 54, Appl
39	84	4.6	791	1	US-08-643-219-1	Sequence 1, Appli
40	84	4.6	791	2	US-09-131-995-1	Sequence 1, Appli
41	84	4.6	791	2	US-08-832-087B-1	Sequence 1, Appli
42	84	4.6	791	3	US-08-851-350-1	Sequence 1, Appli
43	84	4.6	791	4	US-09-132-154-1	Sequence 1, Appli
44	84	4.6	810	1	US-07-854-603-2	Sequence 2, Appli
45	84	4.6	810	1	US-08-147-000B-29	Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-08-933-750C-26
; Sequence 26, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNOBNOT03

; CLONE: 1573677
US-08-933-750C-26

Query Match 100.0%; Score 1818; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.3e-173;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLGGLGCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
DB 1 MRLGGLGCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60

QY 61 HGGQVGLKYSRRAOPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
DB 61 HGGQVGLKYSRRAOPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120

QY 121 NSPGNVQPNAPSVEHPVLEKLAHAHSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
DB 121 NSPGNVQPNAPSVEHPVLEKLAHAHSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180

QY 181 SIWCSTEHNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQD 240
DB 181 SIWCSTEHNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQD 240

QY 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLIISYKH 300
DB 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLIISYKH 300

QY 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340
DB 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340

RESULT 2

US-09-234-613-26
; Sequence 26, Application US/09234613
; Patent No. 6132973

GENERAL INFORMATION:

; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:

CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166
; TELEX:

INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNCLOT03
; CLONE: 1573677
; US-09-234-613-26

Query Match 100.0%; Score 1818; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.3e-173;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLGGLGCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60
DB 1 MRLGGLGCTLRFTSPPTDSSVTETIILCTMLFLGSLGAWGTTSTGSIIFSLKTLRSQ 60

QY 61 HGGQVGLKYSRRAOPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120
DB 61 HGGQVGLKYSRRAOPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGSGGAGSDS 120

QY 121 NSPGNVQPNAPSVEHPVLEKLAHAHSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180
DB 121 NSPGNVQPNAPSVEHPVLEKLAHAHSYNPKFEFNNLKSGRVFIKSYSEDDIHSIKY 180

QY 181 SIWCSTEHNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQD 240
DB 181 SIWCSTEHNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCGVAEMKSPVDYGTSGVWSQD 240

QY 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLIISYKH 300
DB 241 KWKGFQVQWIFVKDVPNNQLRHIRELNDNKPVTNSRDTQEVPLEKAKQVLIISYKH 300

QY 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340
DB 301 TTSIFDDFAHYEKRRRRWCARNRGRVETNNEGEPVSYMF 340

RESULT 3

US-09-269-731-2
; Sequence 2, Application US/09269731
; Patent No. 6333185

GENERAL INFORMATION:

; APPLICANT: BARBEYRON, Tristan
; APPLICANT: POTIN, Philippe
; APPLICANT: RICHARD, Christophe
; APPLICANT: HENRISSAT, Bernard
; APPLICANT: YVIN, Jean-Claude
; APPLICANT: KLOAREG, Bernard

TITLE OF INVENTION: Glycolyse hydrolase genes and their

; TITLE OF INVENTION: use for producing enzymes for the biodegradation of
; TITLE OF INVENTION: carrageenans

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

; ADDRESSEE: DENNISON, MESEROLE, SCHEINER & SCHULTZ
; STREET: 612 Crystal Square 4, 1745 Jefferson Davis
; STREET: Highway
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/269,731

Db 307 PATHGOVYTT-----GSGISSTAATPASAGHV-----MSKQOA 341
Qy 75 QPLPAOPPALAOPVQSPQOPQOTRWAPRNRNNAAFQSGGCA---GSDNSPGNVQPNPSA 131
Db 342 PPPPPQPPQAPP---APQAPQPO-AAPQOPAPQPOQAHTLTLSSEPGOSQ-----393
Qy 132 PSVESHVPLEKLAH-----SYNPKFEFNLKSGRVFIKSYSEDDIHRSIKYSIMCST 186
Db 394 ---RTHIKTEGLSFHSEYQOQHSPQJAYS-----PFLPHYSPPS---YPPITRSQIDYT 443
Qy 187 EHGKNR-----LDSAFRCMS-SKGPVY 207
Db 444 DHQNSSSYSHAAGQGTGLYSTFTYMNPAQRPMY 477

RESULT 8

US-08-586-165-7
; Sequence 7, Application US/08586165
; Patent No. 6054298
; GENERAL INFORMATION:
; APPLICANT: Laufer, Edward M.
; APPLICANT: Orozco, Olivia E.
; APPLICANT: Tabin, Clifford J.
; TITLE OF INVENTION: Fringe Proteins and Pattern Formation
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,165
; FILING DATE: 16-JAN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: H095-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-586-165-7

Query Match 5.08; Score 90; DB 3; Length 363;
Best Local Similarity 20.68; Pred. No. 0.87;
Matches 78; Conservative 54; Mismatches 142; Indels 104; Gaps 19;

Qy 22 SVTETIILCTMLFL-----GSLG-AWGTTSTGTSFSLKTLRSQHGQVGL-----K 68
Db 12 SLVGSMTCLLVLMVEPPGRGLARGEAGGAORALOSLGAARACQAGPLRTADTEGR 71
Qy 69 VSRPRAQPLAOPPALAOPVQSPQOPQ-----TRWV 101
Db 72 LSRARRE-LPAAPP-----SPRPPPAEDITPRDVFIAVKTTKKFKHAKLELLLDTWI 122
Qy 102 APRNRNAAFGSGAGSD-SNSPGNVQPNAPSAPSVESHVPLEKLAHYSNPKFEWNLKS 160
Db 123 S-RNRDMTFITDGEDELKQKQARNVINTNC-SAAHSROALSKMAV-----EYDKFTES 175

Qy 161 GRVFIKSYSEDDIHRSIK---YSIWCSTEH-----GNKRLDSAPRCM-----SSKGPVY 207
Db 176 GRKWF--CHVDDDNVNVVTLVKLLSSYPHTQDIYIGRPSLDRPIQATERISENKMHPVH 233
Qy 208 LLFSVNSGSHFC---GVAEMKSPVDYG-----TSAGVNSQDKWKGFQVQWIFVKDVPNN 259
Db 234 FWFATGGAG-FCISRGLALKMSPWASGGHFMSTAEBKIRLPDDCTIGYIIESVLGVKLIRS 292
Qy 260 QLRHRLNNDNKPVTNSRDTQEVPLEKAKQV---LKIISYKHTTSTFDDFAHVEKQR 316
Db 293 NLFHSHLENLHOVPKT-----EIHQVTLISYGFENKRNKSIHMKGAFSVEEDPSR 342
Qy 317 RRR-----WCARN 324
Db 343 FRSVHCLLYPDTWPCPSN 360

RESULT 9

US-08-718-738-2
; Sequence 2, Application US/08718738
; Patent No. 6013469
; GENERAL INFORMATION:
; APPLICANT: KUNSCH, CHARLES A
; APPLICANT: CHOPRA, ARVIND
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN B-CELL TRANSLOCATION GENES-2 AND 3
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,738
; FILING DATE: 18-SEP-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488-0730003/EKS/KMT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-718-738-2

Query Match 4.88; Score 88; DB 3; Length 345;
Best Local Similarity 27.28; Pred. No. 1.3;
Matches 31; Conservative 16; Mismatches 59; Indels 8; Gaps 4;

Qy 46 ISTGSIFSLKTLRSQHGQVGLKV---SRPRAQPLPAQPPALAQYQSPQPPQTRWA 102
Db 211 LNVNDLLKQAKAISSMHSLYGLGLGSGQQQQAQPPPPPPPPQOQOQO---RTSALS 268
Qy 103 PRNRNAAFGSGAGSDNS--PGNVQPNAPSAPSVESHVPLEKLAHYSNPKF 154
Db 269 PNAKEFITPNQOGSGSSTNGMFGDSPLNLSPLQYSN-AFDVFAAYGGLNEKSF 321

```
RESULT 10
US-09-221-844-2
; Sequence 2, Application US/09221844
; Patent No. 6258777
; GENERAL INFORMATION:
; APPLICANT: KUNSCH, CHARLES A
; APPLICANT: CHOPRA, ARVIND
; APPLICANT: ROSEN, CRAIG A
; TITLE OF INVENTION: HUMAN B-CELL TRANSLOCATION GENES-2 AND 3
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,844
; FILING DATE: 29-DEC-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/718,738
; FILING DATE: 18-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,382
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/460,104
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03323
; FILING DATE: 17-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLONIG, ROBERT C
; REGISTRATION NUMBER: 34,395
; REFERENCE/DOCKET NUMBER: 1488.0730004/EKS/RCM/SCW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-221-844-2

Query Match 4.8%; Score 88; DB 4; Length 345;
Best Local Similarity 27.2%; Pred. No. 1.3;
Matches 31; Conservative 16; Mismatches 59; Indels 8; Gaps 4;

QY 46 ISFGSIFSLKTLRSQHGQVGLKV---SRPRAQLPLPAQPPALAAQYQSPQPPOTRWVA 102
Db 211 LNVNDLLKOKATSSMHSYGLGSGQQPQQQQQPAQPPPPPPPPPPPPPPPPPPPPPP 268
QY 103 PRNRNAAFQSGGAGSDSNS--PGNVQPSAPSHPVLEKLAHSAHYNPKF 154
Db 269 PNAKEFIFFNMQGSSTNGMFGDPLNLSPLQYSN-AFDVFAAYGGLNEKSF 321

RESULT 11
US-08-906-865-1
; Sequence 1, Application US/08906865
; Patent No. 6040168
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
```

```
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,865
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc = "Synapsin III"
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-906-865-1

Query Match 4.8%; Score 87.5; DB 3; Length 580;
Best Local Similarity 30.0%; Pred. No. 3.1;
Matches 30; Conservative 18; Mismatches 49; Indels 3; Gaps 3;

QY 54 LKTLRSQHGQVGLKYSRPRAPQLPAQPPALAAQ-PQYQSPQPPQTRWVAPRNRNAAFQ 112
Db 414 IKSAGSPGAQLGLPQLXQPPRPPPGGPRQAQSPQPSQQR-LSPPGQQPLSPQ 472
QY 113 SGGAGSDSNSPGNVQPSAPSHPVLEKLAHSAHYNPK 152
Db 473 S-GSPQQRSPGSPQLSRASSGSSPNOAKPGATLASOPR 511

RESULT 12
US-08-682-847-4
; Sequence 4, Application US/08682847
; Patent No. 5858989
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE
; APPLICANT: VAN DEN HURK, SYLVIA
; APPLICANT: ZAMB, TIM
; APPLICANT: FITZPATRICK, DAVID
; TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 19:28:11 ; Search time 33.33 Seconds
(without alignments)
1133.066 Million cell updates/sec

Title: US-09-877-633-1
Perfect score: 1818
Sequence: 1 MRLRGLLQGLRFTSPPTD.....CARGRVETNNEGPVSYMF 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length DB	ID	Description	
1	991.5	54.5	570	22	AA139510	Human polypeptide
2	978.5	53.8	564	22	AA141296	Human polypeptide
3	978.5	53.8	579	22	AA136626	Human FLEXIT-48 pr
4	673.5	37.0	700	22	AB162329	Drosophila melanog
5	521	28.7	429	21	AA142527	Arabidopsis thalia
6	521	28.7	511	21	AA142526	Arabidopsis thalia
7	521	28.7	530	21	AA142525	Arabidopsis thalia
8	516	28.4	315	21	AA108665	Arabidopsis thalia
9	515	28.3	257	21	AA146038	Arabidopsis thalia
10	515	28.3	260	21	AA108667	Arabidopsis thalia
11	515	28.3	286	21	AA146037	Arabidopsis thalia

12	515	28.3	289	21	AA108666	Arabidopsis thalia
13	507	27.9	292	21	AA121588	Arabidopsis thalia
14	507	27.9	293	21	AA139341	Arabidopsis thalia
15	507	27.9	294	21	AA121587	Arabidopsis thalia
16	507	27.9	295	21	AA139340	Arabidopsis thalia
17	506.5	27.9	274	21	AA121589	Arabidopsis thalia
18	506.5	27.9	275	21	AA139342	Arabidopsis thalia
19	504	27.7	461	21	AA104736	Arabidopsis thalia
20	504	27.7	465	21	AA104735	Arabidopsis thalia
21	504	27.7	503	21	AA104734	Arabidopsis thalia
22	502	27.6	529	21	AA120629	Arabidopsis thalia
23	502	27.6	1421	21	AA139649	Arabidopsis thalia
24	502	27.6	1499	21	AA139648	Arabidopsis thalia
25	502	27.6	1655	21	AA139647	Arabidopsis thalia
26	496	27.3	493	21	AA120630	Arabidopsis thalia
27	495.5	27.3	471	21	AA120631	Arabidopsis thalia
28	438.5	24.1	326	21	AA113961	Arabidopsis thalia
29	438.5	24.1	355	21	AA113960	Arabidopsis thalia
30	438.5	24.1	388	21	AA113959	Arabidopsis thalia
31	339	18.6	277	21	AA137877	Arabidopsis thalia
32	277	15.2	256	21	AA137879	Arabidopsis thalia
33	217	11.9	45	22	AB111565	Human NY-REN-2 Ag
34	197.5	10.9	407	20	AA107049	Renal cancer assoc
35	196.5	10.8	722	22	AB158524	Drosophila melanog
36	187.5	10.3	712	18	AA130749	Rat YTS21 gene pro
37	179	9.8	274	21	AA137878	Arabidopsis thalia
38	131	7.2	312	22	AB111584	Human NY-REN-2 Ag
39	105	5.8	480	21	AA182317	Human protein tran
40	103.5	5.7	1766	22	AB158631	Drosophila melanog
41	103	5.7	839	21	AA156884	Human prostate can
42	101	5.6	578	22	AA173862	Human colon cancer
43	99	5.4	559	19	AA150909	Alteromonas fortis
44	98.5	5.4	786	22	AB106261	Novel human diagno
45	98.5	5.4	786	22	AB106844	Novel human diagno

ALIGNMENTS

RESULT 1	
AA139510	AA139510 standard; Protein; 570 AA.
XX	AA139510;
AC	AA139510;
DT	22-OCT-2001 (first entry)
XX	Human polypeptide SEQ ID NO 2655.
DE	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-052317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58666.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2655; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 570 AA;

Query Match 54.5%; Score 991.5; DB 22; Length 570;
Best Local Similarity 75.3%; Pred. No. 2.2e-89;
Matches 189; Conservative 19; Mismatches 36; Indels 7; Gaps 4;
QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGAGSDNSPGNVQ--PNS 130
DB 325 qtqplpppppppqalqvqqaaqp-trwvaprnrsgfghn---gvdngvgsgqsgsgs 380
QY 131 APSVESHVPVLEKLAHNSYNPKFEWNLSGSRVFIKSYSEDDIHRSIKYSIWCSTEHGN 190
DB 381 tps-ephvpvleklrsinnynpkdfwnlkhgrvfliksyseddihrsikynlwcstehgn 439
QY 191 KRLSAFCRMSKGPVYLLFVSGHFCGVAEMKSPVDYGTFSAGVSDQDKWKGFDVQW 250
DB 440 krldaayrsmgkpgpyllfsvngsgfhfcgvaemksavdyntcagvwsqdkwkgfdvwr 499
QY 251 IFVKDVPNNQLRHIRLENNNDKNPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
DB 500 ifvkdvpnsqrlhrilennnenkpvtnsrdtqevplekakqvlkllasykhttsifddfh 559
QY 311 YEKRRRRRWC 321
DB 560 yekrrgrkrk 570

RESULT 2
AAW41296
ID AAW41296 standard; Protein; 564 AA.
XX
XX AAW41296;
XX
XX 22-OCT-2001 (first entry)
DT
XX Human polypeptide SEQ ID NO 6227.
DE
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS Homo sapiens.
XX WO200153312-A1.
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60452.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6227; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 564 AA;

Query Match 53.8%; Score 978.5; DB 22; Length 564;
Best Local Similarity 75.9%; Pred. No. 4.2e-88;
Matches 186; Conservative 19; Mismatches 33; Indels 7; Gaps 4;
QY 73 RAQPLPAQPPALAAQPOYQSPQPPQTRWVAPRNRNAAFGQSGAGSDNSPGNVQ--PNS 130
DB 325 qtqplpppppppqalqvqqaaqp-trwvaprnrsgfghn---gvdngvgsgqsgsgs 380
QY 131 APSVESHVPVLEKLAHNSYNPKFEWNLSGSRVFIKSYSEDDIHRSIKYSIWCSTEHGN 190
DB 381 tps-ephvpvleklrsinnynpkdfwnlkhgrvfliksyseddihrsikynlwcstehgn 439
QY 191 KRLSAFCRMSKGPVYLLFVSGHFCGVAEMKSPVDYGTFSAGVSDQDKWKGFDVQW 250
DB 440 krldaayrsmgkpgpyllfsvngsgfhfcgvaemksavdyntcagvwsqdkwkgfdvwr 499
QY 251 IFVKDVPNNQLRHIRLENNNDKNPVTNSRDTQEVPLEKAKQVLKIISSYKHTTSIFDDFAH 310
DB 500 ifvkdvpnsqrlhrilennnenkpvtnsrdtqevplekakqvlkllasykhttsifddfh 559


```
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 700 AA;
    Query Match          37.0%; Score 673.5; DB 22; Length 700;
    Best Local Similarity 48.7%; Pred No. 1.1e-57;
    Matches 150; Conservative 35; Mismatches 84; Indels 39; Gaps 9;
QY 70 SRPRAQPL-----PAQPPALAPQ----YOSPPQPPQTRW-----VAPRRNNAFGQS 113
    || : | || || || || || || || || || || || || || || || || || || ||
Db 260 stakvsptgvrnlgpqpvhaaapratnagpfgppnarthdggphpsrnsersgmys 319
QY 114 GGAGS-----DSNSPGNVQPNAPS---VESHVPVLEKLKAHSYNKPEFEWNLK 159
    || || || || || || || || || || || || || || || || || || || || ||
Db 320 sfirgefesaskfeyrdenqsrpveatsateelpvdsqilvldelkdnynpkvid--lk 377
QY 160 ---SGRVFIKSYSEDIIHRSIKYSIMCSTEGHGNKRLDSAFRCMSSK-GPVYLLFSVNGS 215
    || || || || || || || || || || || || || || || || || || || || ||
Db 378 kagsarffvikyseddhrskyelwctdngnkrlddafkerheeggnlmffsvngs 437
QY 216 GHFCGVAEMKSPVDYGTSGAVMSQDKWKGKFDVQWIEFVKDVPNNQLRHIRLENNDKKPYT 275
    || || || || || || || || || || || || || || || || || || || || ||
Db 438 ghfcgmaqmtpvdynstssvsgdqkrgkfkwklyvkdvpngtlrhirlennnksvt 497
QY 276 NSRDQEVPLEKAKQVLKIISYKHTTSIFDPFAHYEKQRRRWCAR---NGRVETNNE 332
    || || || || || || || || || || || || || || || || || || || || ||
Db 498 nardtqevpdkgievligilhsynhstsfidffhyekkgveevsksrppmhgpdgnnha 557
QY 333 GEPVSYMF 340
    || : |
Db 558 papiarsf 565
RESULT 5
ID AAG42527 standard; Protein; 429 AA.
XX AC AAG42527;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 53048.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 09-MAR-1999; 99US-0123180.
XX PR 23-MAR-1999; 99US-0123548.
XX PR 25-MAR-1999; 99US-0125788.
XX PR 29-MAR-1999; 99US-0126264.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
```


PR	20-JUL-1999;	99JUS-0144884;
PR	21-JUL-1999;	99JUS-0144814;
PR	21-JUL-1999;	99JUS-0145086;
PR	21-JUL-1999;	99JUS-0145088;
PR	22-JUL-1999;	99JUS-0145085;
PR	22-JUL-1999;	99JUS-0145087;
PR	22-JUL-1999;	99JUS-0145089;
PR	22-JUL-1999;	99JUS-0145192;
PR	23-JUL-1999;	99JUS-0145145;
PR	23-JUL-1999;	99JUS-0145218;
PR	23-JUL-1999;	99JUS-0145224;
PR	26-JUL-1999;	99JUS-0145276;
PR	27-JUL-1999;	99JUS-0145913;
PR	27-JUL-1999;	99JUS-0145918;
PR	27-JUL-1999;	99JUS-0145919;
PR	28-AUG-1999;	99JUS-0145951;
PR	28-AUG-1999;	99JUS-0146386;
PR	02-AUG-1999;	99JUS-0146388;
PR	02-AUG-1999;	99JUS-0146389;
PR	03-AUG-1999;	99JUS-0147038;
PR	04-AUG-1999;	99JUS-0147204;
PR	04-AUG-1999;	99JUS-0147302;
PR	05-AUG-1999;	99JUS-0147192;
PR	05-AUG-1999;	99JUS-0147260;
PR	06-AUG-1999;	99JUS-0147303;
PR	06-AUG-1999;	99JUS-0147416;
PR	08-AUG-1999;	99JUS-0147493;
PR	09-AUG-1999;	99JUS-0147935;
PR	10-AUG-1999;	99JUS-0148171;
PR	11-AUG-1999;	99JUS-0148319;
PR	12-AUG-1999;	99JUS-0148341;
PR	13-AUG-1999;	99JUS-0148565;
PR	13-AUG-1999;	99JUS-0148684;
PR	16-AUG-1999;	99JUS-0149368;
PR	17-AUG-1999;	99JUS-0149175;
PR	18-AUG-1999;	99JUS-0149426;
PR	20-AUG-1999;	99JUS-0149722;
PR	20-AUG-1999;	99JUS-0149723;
PR	20-AUG-1999;	99JUS-0149929;
PR	23-AUG-1999;	99JUS-0149902;
PR	23-AUG-1999;	99JUS-0149930;
PR	25-AUG-1999;	99JUS-0150566;
PR	26-AUG-1999;	99JUS-0150884;
PR	27-AUG-1999;	99JUS-0151065;
PR	27-AUG-1999;	99JUS-0151066;
PR	27-AUG-1999;	99JUS-0151080;
PR	30-AUG-1999;	99JUS-0151303;
PR	31-AUG-1999;	99JUS-0151438;
PR	01-SEP-1999;	99JUS-0151930;
PR	22-SEP-1999;	99JUS-0154779;
PR	22-SEP-1999;	99JUS-0155139;
PR	10-SEP-1999;	99JUS-0152363;
PR	10-SEP-1999;	99JUS-0153070;
PR	13-SEP-1999;	99JUS-0153758;
PR	15-SEP-1999;	99JUS-0154018;
PR	16-SEP-1999;	99JUS-0154039;
PR	20-SEP-1999;	99JUS-0154779;
PR	22-SEP-1999;	99JUS-0155139;
PR	23-SEP-1999;	99JUS-0155486;
PR	24-SEP-1999;	99JUS-0155659;
PR	28-SEP-1999;	99JUS-0156458;
PR	29-OCT-1999;	99JUS-0156596;
PR	04-OCT-1999;	99JUS-0157111;
PR	05-OCT-1999;	99JUS-0157753;
PR	06-OCT-1999;	99JUS-0157865;
PR	07-OCT-1999;	99JUS-0158029;
PR	08-OCT-1999;	99JUS-0158232;
PR	12-OCT-1999;	99JUS-0158369;
PR	13-OCT-1999;	99JUS-0159293;
PR	13-OCT-1999;	99JUS-0159294;
PR	13-OCT-1999;	99JUS-0159295;
PR	14-OCT-1999;	99JUS-0159329;
PR	14-OCT-1999;	99JUS-0159330;
PR	14-OCT-1999;	99JUS-0159637;

PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	23-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match		28.7%;	Score 521;	DB 21;	Length 429;
Best Local Similarity		43.8%;	Pred. No. 7.6e-43;		
Matches 125;	Conservative 28;	Mismatches 94;	Indels 40;	Gaps	
Qy	61 HGGVGGLKVSRAQPLPAOPALAO-	-OYQSPQOP-	-----	PQTRWVAPNR--	106
Db	107 hgmddrpktpkraqsqsyappllnq	egkriaypmdpvkksqalnrdeteka	rtken	166	
Qy	107 -----NAAFGQ----	SGGAGSDS--NSPGNVQPSAPS	VESHVPLEKLAHAHSYNPKFEW	156	
Db	167 gtsmdliangdhtingecescsida	egnersngvsvirr-----	dqynlpsfq	217	
Qy	157 NLKSGRVFIKSYEDDIHSIKYSI	WCSTEHGNKRLDSAPR-----	CMSSKGPVYLL	209	
Db	218 kyeaifviksydsedihksikynv	wsstlngnkkldsaygesqkaadksg	kcpvfif	277	
Qy	210 FSVNGSHFCGVAEMKSPVDYGT	SAGVMSODKWKGFVOWIFVKDVP	NNOLRHIRLENN	269	
Db	278 fsvnasqfcvqaemigrvdyeksm	efwqddkwtgyfpwkwhiikvnpq	lrhiilenn	337	
Qy	270 DNKPVTNSRDTQEVPLEKAKQV	LIKTISSYKHTTSIFDDFAHYEK	RQR	316	
Db	338 enkpvtnsrtdqevrlpqgnev	lnifknyaaktsildddfayenrek	384		

RESULT 6	
AA	AG42526
ID	AG42526 standard; Protein; 511 AA.
AC	AG42526;
XX	18-OCT-2000 (first entry)
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 53047.
DE	Protein identification; signal transduction pathway; metabolic pathwa
KW	hybridisation assay; genetic mapping; gene expression control; promot
KW	termination sequence.
XX	Arabidopsis thaliana.
XX	EPI033405-A2.
XX	06-SEP-2000.
XX	25-FEB-2000; 2000EP-0301439.
XX	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	18-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 28.7%; Score 521; DB 21; Length 511;			
Best Local Similarity 43.6%; Pred. No. 9.9e-43;			
Matches 125; Conservative 28; Mismatches 94; Indels 40; Gaps 7;			
QY	61	HGGVGLKVRPRAPPLPAQPPALAAQP--QYQSPQQP-----PQTRWVAPRNR--	106
Db	189	hgmgdrptkprkasqnsyappplnqkgriaypmdpvkksgalnrdekekakartken	248
QY	107	-----NAAFGO----SGAGSDS--NSPGNVQPNAPSVESHPVLEKLUKAHSYNPKFEFW	156
Db	249	gtsmndlangqdhntgcecscldaegnersngvsvir-----dqynlpsfqt	299
QY	157	NLKSGRVFIITKSYSEDDIHSIKYISIMCSTEHGNKRLDLSAFR-----CMSSKGPVYLL	209
Db	300	kyeaaifvikiyseddihsikynvwsstlngnkkidsayqesqkkaadskcpvflf	359
QY	210	FSVNGSCHFCGVAEMKSPVDYGTSGAGVSDQWKKGKFDVQWIFVKDVPNNQLRHIRENN	269
Db	360	fsvnasgfcgvaemigrvdyeksmefwqgdkwtgyfpvkwhiikdvnpqlrhiilenn	419
QY	270	DNKPVTVNSRDTQVPLEKAKQVLKLISSYKHTTSIFDDFAHYEKROR	316
Db	420	enkpvtnsrdtqvrlpqgnevlnifknyaaktsilddffeyenrek	466
RESULT 7			
AAG42525			
ID	AAG42525 standard; Protein; 530 AA.		
XX	AC	AAG42525;	
XX	DT	18-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 53046.	
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;	
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	OS	termination sequence.	
XX	PN	Arabidopsis thaliana.	
XX	PN	EPI033405-A2.	

XX	06-SEP-2000.	
PD		
XX		
PE	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 6294.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.

PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 28.4%; Score 516; DB 21; Length 315; Best Local Similarity 44.0%; Pred. No. 1.5e-42; Matches 122; Conservative 28; Mismatches 87; Indels 40; Gaps			
QY	71 RPRAPLPAQPPALAAQP--QYQSPQOP-----PQTRWVAPNR-----NAAFG 111		
Db	3 rkasqsyapppllndegkriaypmdpvkkgalnrdekakartkengtsmdldang 62		
QY	112 Q-----SCGAGSDS-NSPGNVQPNAPSVESHVPVLEKLKAASHYNPKFEFNNLKSGRVFI 166		
Db	63 qdhitngecescsldaeqnersngvsvrr-----dqynlpsftkkyeeaffvl 113		
QY	167 KYSSEDDIHRSTIKYSTWCSTHGKRLDSAFR-----CMSSKGPVYLLFSVNGSGHFC 219		
Db	114 ksyseddihkstkynwsstlgnkklidsayqesqkkaadskgkcpvflffsvnasgqfc 173		
QY	220 GVAEMKSPVDYGTSGAGVMSQDKWKFQVQWIFVKDVPNNQLRHIRELNNDKRPVTNSRD 279		
Db	174 gvaemigrdyeksmefwqdgkwtgfyfvkwhiikdvpnpqrpqhiilenenkpvtnsrd 233		
QY	280 TOEVRLEKAKQVLKIISSTYKHTTSIFDDFAHYEKOR 316		
Db	234 tqevrlpqgnevlnifknyaaktslliddfdyenrek 270		

RESULT	9
AAG46038	
ID	AAG46038 standard; Protein; 257 AA.
XX	
AC	AAG46038;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 57874.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	termination sequence.
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.

PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149920;
PR	23-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	25-AUG-1999;	99US-0150566;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	30-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151438;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153750;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155133;
PR	22-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	06-OCT-1999;	99US-0157753;
PR	07-OCT-1999;	99US-0157865;
PR	08-OCT-1999;	99US-0158029;
PR	12-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

2:

Qy	121	NSPGNVOPNSAPS	VESHVPLEK	LKAHSHYNPK	EFENWLKSGR	VFIIKSYSEDD	IHSIKY	180
		: :	:	:	:	:	:	
Db	22	daegnersngvgvirr	-----	-dqynlpsfqkyeeaf	ifvfkysredd	ihksiky	72	
		: :	:	:	:	:	:	
Qy	181	SIWCSTEHGNKR	LDSAPR	-----	-CMSSGKGPVYL	LFSSVNGSHGFCV	APKSPVDYGYFS	233
		: :	:	:	:	:	:	
Db	73	nwsstingnkkidsa	ygesqkkaadskg	gkcpvlf	ffsvnasgqf	cgvaemigrvdeks	132	
		: :	:	:	:	:	:	
Qy	234	AGVWSQDKWKGK	FDVQWIFVKD	VPNNQLRH	IRLENDNKKPVT	NSRDTQEVPLEKA	QVULK	2933
		: :	:	:	:	:	:	
Db	133	mefwgqgdwktay	fvykwhiikd	vgpncprr	lilennenkpvtrsd	rtqevrlpqaaevln	192	
		: :	:	:	:	:	:	

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 28.3%; Score 515; DB 21; Length 286;
Best Local Similarity 51.7%; Pred. No. 1,6e-42;
Matches 105; Conservative 22; Mismatches 60; Indels 16; Gaps 2;
QY 121 NSFGNVQPNAPSVEHPVLEKLAHNSYNPKFEWNLKSGRVFIITKSYSEDDIHRSIKY 180

```
Db 51 daegnersngvgsvrr-----dqynlpsftktyeeaaiffviksyseddihksiky 101
Qy 181 SIWCTEHGKNRLDSAFR-----CMSSKGPVYLLFSVNGSGHFHCGVAEMKSPVDYGTG 233
Db 102 nvsvstingnkkldsaygesqkkaadskgkcpvflfsvnasgqfcgvaemigrvdyeks 161
Qy 234 AGVNSQDKWKGKFDVQWIFVKDVPNNOLRHIRLENNDKPVTNSRDTQEVPLEKAKQVLK 293
Db 162 mefwqdkwtgyfvpkwhiikdvnpqlrhiilennenkpvtnsrtdqevrlpqagnevln 221
Qy 294 IISYKHTTSIFDDFAHYERQR 316
Db 222 ifknyaaktsildddfyenrek 244

RESULT 12
AAG08666
ID AAG08666 standard; Protein; 289 AA.
XX
AC AAG08666;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6295.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 28-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
```

PR	05-AUG-1999;	99US-0147192.	Query Match	28.3%;	Score 515;	DB 21;	Length 289;
PR	06-AUG-1999;	99US-0147260.	Best Local Similarity	51.7%;	Pred. No. 1.6e-42;		
PR	06-AUG-1999;	99US-0147303.	Matches 105;	Conservative 22;	Mismatches 60;	Indels 16;	Gaps 2;
PR	09-AUG-1999;	99US-0147416.					
PR	09-AUG-1999;	99US-0147493.					
PR	09-AUG-1999;	99US-0147935.					
PR	10-AUG-1999;	99US-0148171.					
PR	11-AUG-1999;	99US-0148319.					
PR	12-AUG-1999;	99US-0148341.					
PR	13-AUG-1999;	99US-0148565.					
PR	13-AUG-1999;	99US-0148684.					
PR	16-AUG-1999;	99US-0149368.					
PR	17-AUG-1999;	99US-0149175.					
PR	18-AUG-1999;	99US-0149426.					
PR	20-AUG-1999;	99US-0149722.					
PR	20-AUG-1999;	99US-0149723.					
PR	20-AUG-1999;	99US-0149929.					
PR	23-AUG-1999;	99US-0149902.					
PR	23-AUG-1999;	99US-0149930.					
PR	25-AUG-1999;	99US-0150566.					
PR	26-AUG-1999;	99US-0150884.					
PR	27-AUG-1999;	99US-0151065.					
PR	27-AUG-1999;	99US-0151066.					
PR	27-AUG-1999;	99US-0151080.					
PR	30-AUG-1999;	99US-0151303.					
PR	31-AUG-1999;	99US-0151438.					
PR	01-SEP-1999;	99US-0151930.					
PR	07-SEP-1999;	99US-0152363.					
PR	10-SEP-1999;	99US-0153070.					
PR	13-SEP-1999;	99US-0153758.					
PR	15-SEP-1999;	99US-0154018.					
PR	16-SEP-1999;	99US-0154039.					
PR	20-SEP-1999;	99US-0154779.					
PR	22-SEP-1999;	99US-0155139.					
PR	23-SEP-1999;	99US-0155486.					
PR	24-SEP-1999;	99US-0155639.					
PR	28-SEP-1999;	99US-0156458.					
PR	04-OCT-1999;	99US-0157117.					
PR	05-OCT-1999;	99US-0157753.					
PR	06-OCT-1999;	99US-0157865.					
PR	07-OCT-1999;	99US-0158029.					
PR	08-OCT-1999;	99US-0158232.					
PR	12-OCT-1999;	99US-0158369.					
PR	13-OCT-1999;	99US-0159293.					
PR	13-OCT-1999;	99US-0159294.					
PR	13-OCT-1999;	99US-0159295.					
PR	14-OCT-1999;	99US-0159329.					
PR	14-OCT-1999;	99US-0159330.					
PR	14-OCT-1999;	99US-0159331.					
PR	14-OCT-1999;	99US-0159637.					
PR	14-OCT-1999;	99US-0159638.					
PR	18-OCT-1999;	99US-0159584.					
PR	21-OCT-1999;	99US-0160741.					
PR	21-OCT-1999;	99US-0160767.					
PR	21-OCT-1999;	99US-0160768.					
PR	21-OCT-1999;	99US-0160770.					
PR	21-OCT-1999;	99US-0160814.					
PR	21-OCT-1999;	99US-0160815.					
PR	22-OCT-1999;	99US-0160980.					
PR	22-OCT-1999;	99US-0160981.					
PR	22-OCT-1999;	99US-0160989.					
PR	25-OCT-1999;	99US-0161404.					
PR	25-OCT-1999;	99US-0161405.					
PR	25-OCT-1999;	99US-0161406.					
PR	26-OCT-1999;	99US-0161359.					
PR	26-OCT-1999;	99US-0161360.					
PR	26-OCT-1999;	99US-0161361.					
PR	28-OCT-1999;	99US-0161920.					
PR	28-OCT-1999;	99US-0161992.					
PR	28-OCT-1999;	99US-0161993.					
PR	29-OCT-1999;	99US-0162142.					
QY	121 NSFGNVQPNAPSVPVLEKLAHAHSYNPKFEPENLKSGRVFIKISYSEDDIHSIKY 180						
DB	51 daegnersngvsgvirr-----dqynlpsfqkyeeaffvixyseddhksiky 101						
QY	181 SJWCSTEHNKRLDSAFR-----CMSSKGPVYLLFSVNGSCHFCGVAEMKSPVDYGT 233						
DB	102 nwsstclngnkklidsayqesqkkaadksqcpvlfsvnasggqfcgvaeimgrvdyeks 161						
QY	234 AGVWSODKWKGFQVQWIFVKDVPNNQLRHIRLENNNDKRPVTNSRDTQVPLEKAKQVLIK 293						
DB	162 mefwqdkwtgyfvpkwhlkdvppqlrhilennenkpvtnstdtqevrlpqgnevln 221						
QY	294 IISYKHHTSIFDDFAHYEKQR 316						
DB	222 lknyaaktsilddfdyenrek 244						
RESULT 13							
AAG21588							
ID	AAG21588 standard; Protein; 292 AA.						
XX							
AC	AAG21588;						
XX							
DT	17-OCT-2000 (first entry)						
XX							
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 24198.						
XX							
KW	Protein identification; signal transduction pathway; metabolic pathway;						
KW	hybridisation assay; genetic mapping; gene expression control; promoter;						
XX	termination sequence.						
OS	Arabidopsis thaliana.						
PN	EP1033405-A2.						
PD	06-SEP-2000.						
XX							
PF	25-FEB-2000; 2000EP-0301439.						
XX							
PR	25-FEB-1999;	99US-0121825.					
PR	05-MAR-1999;	99US-0123180.					
PR	09-MAR-1999;	99US-0123548.					
PR	23-MAR-1999;	99US-0125788.					
PR	25-MAR-1999;	99US-0126264.					
PR	29-MAR-1999;	99US-0126785.					
PR	01-APR-1999;	99US-0127462.					
PR	06-APR-1999;	99US-0128234.					
PR	08-APR-1999;	99US-0128714.					
PR	16-APR-1999;	99US-0128845.					
PR	19-APR-1999;	99US-0130077.					
PR	21-APR-1999;	99US-0130449.					
PR	23-APR-1999;	99US-0130510.					
PR	28-APR-1999;	99US-0130891.					
PR	30-APR-1999;	99US-0131449.					
PR	30-APR-1999;	99US-0132048.					
PR	30-APR-1999;	99US-0132407.					
PR	04-MAY-1999;	99US-0132484.					
PR	05-MAY-1999;	99US-0132485.					
PR	06-MAY-1999;	99US-0132486.					
PR	07-MAY-1999;	99US-0132487.					
PR	07-MAY-1999;	99US-0132863.					
PR	11-MAY-1999;	99US-0134256.					
PR	14-MAY-1999;	99US-0134218.					
PR	14-MAY-1999;	99US-0134219.					
PR	14-MAY-1999;	99US-0134221.					
PR	14-MAY-1999;	99US-0134370.					
PR	18-MAY-1999;	99US-0134768.					
PR	19-MAY-1999;	99US-0134941.					
PR	20-MAY-1999;	99US-0135124.					

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 16:44:49 ; Search time 237.28 Seconds
(without alignments)
14674.235 Million cell updates/sec

Title: US-09-877-633-2
Perfect score: 2028
Sequence: 1 caaaaggacagataataaa.....agtcacagacagtgtgatgga 2028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1773	87.4	3043	22	AAH72969 Human cervical can
2	1534.6	75.7	2685	22	ABA08628 Human NY-REN-2 Ag
3	870	42.9	8382	22	AAK70731 Human immune haema
4	341.6	16.8	1718	20	AAZ17776 Human gene express
5	341.6	16.8	2750	22	AAK88117 Human FLEXHT-48 nu
6	341.6	16.8	3029	22	AAI60452 Human polynucleoti
7	340.6	16.8	3845	22	AAI58666 Human polynucleoti
8	225	11.1	379	22	AAI66840 Novel human polynu
9	225	11.1	226	14	AAQ61419 Human brain Expres

c	10	218.4	10.8	625	20	AAK90820	Rat progression re
	11	213	10.5	747	22	AAI95660	Human neuroblastom
	12	209.6	10.3	2218	23	ABL11025	Drosophila melanog
c	13	195.4	9.6	6461	23	ABL11024	Drosophila melanog
c	14	173.2	8.5	307	22	AAI08157	Human breast cance
c	15	173.2	8.5	307	22	AAI19067	Human breast cance
c	16	171.4	8.5	293	22	AAI25902	Human breast cance
	17	167.4	8.3	1514	20	AAK39643	Renal cancer assoc
	18	158	7.8	711	20	AAZ15119	Human gene express
	19	144.4	7.1	950	21	AAA44533	Human secreted exp
	20	139.4	6.9	1679	21	AAK32849	Arabidopsis thalia
	21	136.6	6.7	1135	21	AAK39320	Arabidopsis thalia
	22	136.6	6.7	1136	21	AAK46031	Arabidopsis thalia
	23	135.4	6.7	2675	14	AAQ41914	Soybean seed beta-
	24	133.8	6.6	779	22	AAI97361	Human neuroblastom
	25	132.4	6.5	861	21	AAK48546	Arabidopsis thalia
	26	132.4	6.5	1209	21	AAK34339	Arabidopsis thalia
	27	132.4	6.5	1849	21	AAK47231	Arabidopsis thalia
	28	122.8	6.1	739	22	AAI95659	Human neuroblastom
	29	122.2	6.0	729	20	AAV87783	EST clone ES306.
	30	120.2	5.9	1853	21	AAK45480	Arabidopsis thalia
	31	119.4	5.9	1854	21	AAK38976	Arabidopsis thalia
	32	119.4	5.9	4968	21	AAK46149	Arabidopsis thalia
	33	114.8	5.7	925	20	AAV88879	EST clone HZ103.
	34	112	5.5	464	21	AAA42435	Human secreted exp
c	35	103	5.1	400	22	AAK25444	Human ovarian PCR-
	36	101.8	5.0	1586	21	AAK36395	Arabidopsis thalia
	37	94.8	4.7	1749	22	ABA08809	Human renal cancer
	38	94.4	4.7	488	22	ABA09487	Human secreted expre
	39	92	4.5	460	21	AAA43413	Human secreted pro
	40	84.4	4.2	214	21	AAK07592	Human secreted pro
	41	82.4	4.1	246	21	AAK07593	Human secreted pro
	42	66.4	3.3	399	22	AAK67243	Novel human polynu
	43	65.4	3.2	14006	24	ABL33958	Human immune syste
	44	64.8	3.2	3261	22	AAK26571	DNA encoding human
	45	63.8	3.1	6145	24	ABL32972	Human immune syste

ALIGNMENTS

RESULT 1

AAH72969

ID AAH72969 standard; cDNA; 3043 BP.

XX AC AAH72969;

XX AC AAH72969;

DT 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 4243.

DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

XX WO200142467-A2.

PD 14-JUN-2001.

XX 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

DR

XX New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
XX
PS Claim 1; Page 931; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
XX Sequence 3043 BP; 735 A; 781 C; 738 G; 785 T; 4 other;
SQ

Query Match 87.4%; Score 1773; DB 22; Length 3043;
Best Local Similarity 90.0%; Pred. No. 0;
Matches 2027; Conservative 0; Mismatches 0; Indels 224; Gaps 3;

QY 1 caaaaggacaaataaataagatacaaaaatgggttcgtttacatcagaagaagatacagttcatg 60
DB 257 caaaaggacaaataaataagatacaaaaatgggttcgtttacatcagaagaagatacagttcatg 316

QY 61 acaatgaactttagccctaccctactgacagtcacaaatcagagtaacagttaccctctcaa 120
DB 317 acaatgaactttagccctaccctactgacagtcacaaatcagagtaacagttaccctctcaa 376

QY 121 tgagcgaccctacactgtccagctattaccgcgcgtccattggatttcttactccctca 180
DB 377 tgagcgaccctacactgtccagctattaccgcgcgtccattggatttcttactccctca 436

QY 181 atgaggctccggttactgacgggacccctccgattccatcaccaccactacggac 240
DB 437 atgaggctccggttactgacgggacccctccgattccatcaccaccactacggac 496

QY 241 agctcagtaacgagagaccatcatcttctacagatgctgttttggcgacctgggggcc 300
DB 497 agctcagtaacgagagaccatcatcttctacagatgctgttttggcgacctgggggcc 556

QY 301 tggggaaacaatctatcagcagagttcaatttttccctgaaacacctggttctcag 360
DB 557 tggggaaacaatctatcagcagagttcaatttttccctgaaacacctggttctcag 616

QY 361 catgggggacaagtgggttctcaaggtcagcagaccacaga----- 399
DB 617 catgggggacaagtgggttctcaaggtcagcagaccacagaagctccgcgtatgggagcagct 676

QY 400 ----- 399
DB 677 acacctaccctccagctccctgggtggcagcgggtggtgatggcgaccaggtttcaca 736

QY 400 ----- 399
DB 737 gcgacacctcagcaaggccccctccaccataaaagcataaactgagacattggcaactgg 796

QY 400 ----- 399
DB 797 gataacaaggcgctgtgcgaaggccccagttccccccagcaggacccctctccacaggct 856

QY 400 -----gtcagcctctcccagcacagccccacagcttttgggtctcaa 438
DB 857 gccccaacagccccagcaggtggctcagctctcccagcacagccccacagcttttgggtctcaa 916

QY 439 ccgcagtaacagccctcagcagccacccccagaccctgggttgggttccccacgcaacaga 498
DB 917 ccgcagtaacagccctcagcagccacccccagaccctgggttgggttccccacgcaacaga 976

QY 499 aacgcggcgtttgggcagagcggagggtggcgagcgatagcaactctctctggaacgctc 558
DB 977 aacgcggcgtttgggcagagcggagggtggcgagcgatagcaactctctctggaacgctc 1036

QY 559 cagcctaattctccccccagcgtcgaaatccccaccocgctccttgaaaaactgaaggctgct 618
DB 1037 cagcctaattctccccccagcgtcgaaatccccaccocgctccttgaaaaactgaaggctgct 1096

QY 619 cacagctacaacccgaaaagatttgagtggaatctgaaaagcgggctgtgttctcatcctc 678
DB 1097 cacagctacaacccgaaaagatttgagtggaatctgaaaagcgggctgtgttctcatcctc 1156

QY 679 aagagctactctgagagacacatccaccgctccattaaagtactccactctggttagcaca 738
DB 1157 aagagctactctgagagacacatccaccgctccattaaagtactccactctggttagcaca 1216

QY 739 gagcacggcaacaagcgcttgacagcgcttcgcgtgcagtaucaaaaggggcccgtc 798
DB 1217 gagcacggcaacaagcgcttgacagcgcttcgcgtgcagtaucaaaaggggcccgtc 1276

QY 799 tacctgctcttcagcgtcaaatgggagtgggcatttttgggtggccgagatgaugtc 858
DB 1277 tacctgctcttcagcgtcaaatgggagtgggcatttttgggtggccgagatgaugtc 1336

QY 859 ccgctgactacagcaccagtcgcggggtctgctctcagacacaagtggaaagggaagt 918
DB 1337 ccgctgactacagcaccagtcgcggggtctgctctcagacacaagtggaaagggaagt 1396

QY 919 gatgtccagtggatttttggtaaggatgtaccctaaataaccagctcccgacatcaggctg 978
DB 1397 gatgtccagtggatttttggtaaggatgtaccctaaataaccagctcccgacatcaggctg 1456

QY 979 gagaataagacacaacaaocggtcaacaaactccgggacacccagagggtgccttagaa 1038
DB 1457 gagaataagacacaacaaocggtcaacaaactccgggacacccagagggtgccttagaa 1516

QY 1039 aagaccaagcaagtgcgaaattatcagttcctacagacacacaacccctccactctcgac 1098
DB 1517 aagaccaagcaagtgcgaaattatcagttcctacagacacacaacccctccactctcgac 1576

QY 1099 gactttgctcactacagagaagcgcca--gaggagagagagtggtgcgcaaggaaacggcag 1157
DB 1577 gactttgctcactacagagaagcgccagggagagagagagtggtgcgcaaggaaacggcag 1636

QY 1158 agtcgaaaaacaaatgaggggcgaacacagtttcttacaatgtcttaacgctttgactttgaa 1217
DB 1637 agtcgaaaaacaaatgaggggcgaacacagtttcttacaatgtcttaacgctttgactttgaa 1696

QY 1218 aacagtttaaaacacagtgctgtctggtcagctccagtgctgcgccggtcgggggttgag 1277
DB 1697 aacagtttaaaacacagtgctgtctggtcagctccagtgctgcgccggtcgggggttgag 1756

QY 1278 tgttgcatctttgcctttcttctgttctgttgcatttttgcacagatggatctgcatttttgt 1337
DB 1757 tgttgcatctttgcctttcttctgttctgttgcatttttgcacagatggatctgcatttttgt 1816

QY 1338 actttttctatgtattataatcctgttagaagtcactaataaaggagatttttttttgc 1397
DB 1817 actttttctatgtattataatcctgttagaagtcactaataaaggagta-ttttttttgc 1875

QY 1398 agcttcatcagactgatcctaagtgaatgaagtatcccttaaaaaacaaagcatcta 1457
DB 1876 agcttcatcagactgatcctaagtgaatgaagtatcccttaaaaaacaaagcatcta 1935

QY 1458 ttttggcagaaaatgttctttaaattcagctatttgatattctgtgagacttctcatatt 1517
DB 1936 ttttggcagaaaatgttctttaaattcagctatttgatattctgtgagacttctcatatt 1995

QY 1518 ctcatcccttattgtcttttttagcaaacataagaacacctagtcatttttctattttaga 1577
DB 1996 ctcatcccttattgtcttttttagcaaacataagaacacctagtcatttttctattttaga 2055

QY 1578 gtattctgataaactctcttgaaaataactgaaatcaaaaggttaagtattttttgttctcat 1637
DB 2056 gtattctgataaactctcttgaaaataactgaaatcaaaaggttaagtattttttgttctcat 2115

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 25543; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM62170 to AAM91921. (I) have cycostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX cancers and cancer metastases of haematopoietic-related diseases, especially
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
SQ Sequence 8382 BP; 2411 A; 2122 C; 1929 G; 1920 T; 0 other;

Query Match 42.9%; Score 870; DB 22; Length 8382;
Best Local Similarity 99.9%; Pred. No. 1.5e-216;
Matches 881; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1147 aggaacgagcagagtcgaaacaaacaaatgagggcgacacgtttcttaccatgtcttaacgt 1206
Db 1308 AGGAACGGCAGAGTCGAAACAAACAAATGAGGCGAACCACTTCTTACATGTTCTAACGT 1249
QY 1207 ttgactttgaaacagttttaaacaacgtagctgtggtcagctccagtgatgtctcccg 1266
Db 1248 TTGACTTTGAAACACAGTTTAAACACAGTGTGCTTGTGTCAGTCCAGTGTGTCGCCGTG 1189

Db 1696 gccactgcgacacattcctcctagagaaacagagataaaccagtgaccaactctagg 1755
Qy 1016 acaccagaggtgctcttagaaaaagccaaagctgctgaaattatcagttctaca 1075
Db 1756 aacctcaggaagtcctctggaagagctgaagcaggtgctgaaattatcagcagctaca 1815
Qy 1076 agcacaaacctccattcttcagcagactttgctcactacgagaagcgc-agaagagagg 1134
Db 1816 agcacaccattccattttttgatgacttctcacactatgagaacccaagagaagaag 1875
Qy 1135 agtgtgctgcgaaggaacgcgagagtcgaaacaaacaa 1172
Db 1876 aaagtgttaaaaggaagcgtcaaggtcgtgggaataaa 1913

RESULT 6
AAI60452/c
ID AAI60452 standard; cDNA; 3029 BP.
XX
AC AAI60452;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4441.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 05-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dirmnac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM41296.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4441; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 3029 BP; 766 A; 737 C; 715 G; 809 T; 2 other;

Query Match 16.8%; Score 341.6; DB 22; Length 3029;
Best Local Similarity 67.5%; Pred. NO. 1.1e-78;
Matches 512; Conservative 0; Mismatches 239; Indels 7; Gaps 2;
Qy 416 cacagccccagcttggctcaaccagcagtcacagagcctcagcagccaccagacc 475
Db 1593 CTCACCTCCACCACACAGCCTGCCAGCTTTCAGTCCAGCAACAGGCAGCTCAGCCAACCC 1534
Qy 476 gctgggttggccacgcacacagaaacgcggcgtttgggcagagcgaggggctggcagcg 535
Db 1533 GCTGGGTAGCACCTCGGAACCCGTGGCAGTGGGTTCAGTCAATATGGGTGGATGGTAATG 1474
Qy 536 atagcaactctcctggaaacgctccagcctaatctctgccccagcgtcgaatcccaccccg 595
Db 1473 GAGTAGGACAGTCTCAGGCTGGTCTTGGATCTACTCTCTTC-----AGAACCCACCCAG 1420
Qy 596 tccttgaataaacgaaggtgctcacagctcacacccgaagagatttgagtgaaactga 655
Db 1419 TGTGGAGAGAGCTTCGGTCCATTAACTATTAACCCCAAGAGATTTTGACTGGAATCTGA 1360
Qy 656 aaagcggggcgtgtgttcacatcatcaagaagctactctcagagcagacatccacgcctcatta 715
Db 1359 AACATGCCCGGTTTTCATCATTTAAGAGCTACTCTCAGGAGGATATTCACCGTTCCATTA 1300
Qy 716 agtactccatctgtgttagcacagagcaagcgcaacagcgctggacagcgcttcocgct 775
Db 1299 AGTATAATATTTGGTGCAGCACAGAGCATGGTAAACAGAGACTGGATGCTGCTTATCGTT 1240
Qy 776 gcatgagcagcaagggcccgctcctcctcctcctcctcctcctcctcctcctcctcctcct 835
Db 1239 CCATGAACGGGAAGGCCCGCTTACTTACTTTTCACTAGTCAACGGCAGTGGACACTTCT 1180
Qy 836 gtgggggtggccagagtgaggtcccgctggactacgagcagcagcagcagcagcagcagcagc 895
Db 1179 GTGGCGTGGCAGAAATGAAATCTGCTGGACTACAACACATCTGCAGGCTGTGTGGTCCC 1120
Qy 896 aggacaagtggaaagggaggttggatgctcagtgagatttttgaaggtatgtaccacaata 955
Db 1119 AGGACAAATGGAAGGGTCGTTTTGATGTCAGTGGGATTTTTTGTGAAGGAGCTTCCCAATA 1060
Qy 956 accagctccggcacatcaggtgagaaataaacgacacaaacccggtcacaaaccccggg 1015
Db 1059 GCCAATGGACACATTCGCCTAGAGAACCAACGAGATAAACCCAGTGACCACTCTAGGG 1000
Qy 1016 acaccagaggtgccccttagaaaaagccaagcaagtcgctgaaattatcagttcctaca 1075
Db 999 ACACCTCAGGAAGTGCCTCTGGAAAAGGCTAAGCAGGTGTGAAAATATATAGCCAGCTACA 940
Qy 1076 agcacacacctccatctcttcagcagactttgctcactacagagaagcgc-agaagagagg 1134
Db 939 AGCACACCACCTTCCATTTTGTGATGACTTCTCAGACTATCAGAACCCGCAAGAGGAAG 880
Qy 1135 agtgtgctgcgaggaacgcgagagtcgaaacaaacaa 1172
Db 879 AAGTGTAAAAAGGAACGCTCAAGCTCGTGGGAATAA 842
RESULT 7
AAI58666
ID AAI58666 standard; cDNA; 3845 BP.
XX
AC AAI58666;

[illegible]

RESULT 15
AAL19067/C
ID AAL19067 standard; cDNA: 307 BP.
XX
XX
XX AAL19067;
AC
XX

Db 65 GCTTAGAAATAATGACAAACCGGTTACCAATTCACAGGGACACTCAAGAGGTACC 8

Search completed: July 29, 2002, 18:06:59
Job time: 4930 sec

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 14:57:24 ; Search time 1621.09 seconds
(without alignments)
16884.827 Million cell updates/sec

Title: US-09-877-633-2
Perfect score: 2028
Sequence: 1 caaaaggacaagataataaa.....agtcacagacagtgtgatgga 2028

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	674.6	32.3	1109	10 BE894361	BE894361 601432947
2	650.8	32.1	852	10 BE892982	BE892982 601434159
3	648.8	32.0	1014	9 AV079675	AV079675 AV079675
C 4	644.6	31.8	661	9 AL574115	AL574115 AL574115
C 5	633.4	31.2	1068	9 AL515908	AL515908 AL515908
C 6	628.8	31.0	684	9 AW001012	AW001012 wr91e06.x
7	626.6	30.9	932	9 AV726565	AV726565 AV726565
8	623.8	30.8	1479	10 BM423045	BM423045 PLATE1.D1
C 9	601.8	29.7	664	10 BE646471	BE646471 7e87d05.x
10	594.6	29.3	689	10 BG911605	BG911605 602812729
C 11	583.2	28.8	600	9 AW387914	AW387914 MR4-ST011
12	581.8	28.7	808	10 BE892829	BE892829 601433961
13	580.4	28.6	736	10 BI091481	BI091481 602858960
14	571	28.2	669	9 AW964022	AW964022 EST376095
C 15	555	27.4	582	10 BI262492	BI262492 602953767
16	553.4	27.3	929	10 BG249197	BG249197 602361623
C 17	551	27.2	563	9 AW387761	AW387761 MR4-ST011

ALIGNMENTS

RESULT 1

BE894361
LOCUS BE894361 601432947F1 NIH_MGC_72 1109 bp mRNA linear EST 20-OCT-2000
DEFINITION 601432947F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3918154 5', mRNA sequence.
ACCESSION BE894361
VERSION BE894361.1 GI:10356650
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1109)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTDP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Cloning Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9745 row: n column: 11
High quality sequence stop: 725.
Location/Qualifiers
1. 1109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3918154"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 Kb. Library constructed by Life Technologies."

BASE COUNT 291 a 303 c 273 g 242 t

ORIGIN

Query Match 33.3%; Score 674.6; DB 10; Length 1109;
Best Local Similarity 92.9%; Pred. No. 4e-124;
Matches 763; Conservative 0; Mismatches 49; Indels 9; Gaps 5;

QY 552 aaacgtccagcctaattctgccccagctcgaatccaccctcgtctgaaacactgaa 611
DB 1 AAACGTCACAGCTAAATCTGCCCCAGCTGCAATCCACCCCGCTCTTGAAAACTGAA 60

QY 612 ggcgtctacagctacaaacccgaaagatttgagtggaatctgaaagcggcgtgtgtt 671
DB 61 GGCCTGCTACAGCTACAAACCCGAAAGATTGAGTGGAAATCTGAAAGCGGGCGTGT 120

QY 672 catcatcaagagctactctgagacacacatccaccgc-tccattaaagtaccatcgtt 730
DB 121 CATCATCAAGAGCTACTCTGAGGACGACATCCACCCGCTCCATTAAGTACTCCATCTGGT 180

QY 731 gtacacagagcagcgaaacagcgcctggacagcgccctccgctgcatgagcagcaag 790
DB 181 GTAGCAGACGACGCAACAGCGCTTGACAGCGCTTCCGCTGCATGAGCAGCAAG 240

QY 791 ggcctctacacbtctcttcagcgtcgaatgggagtgggcatttttgggggtggccgaga 850
DB 241 GGCCTGCTACCTGCTCTTCCAGCGTCAATGGGAGTGGCAATTTTGTGGGGTGGCCGAGA 300

QY 851 tgaagtcctcccggtgactcagcagcagcagtcgcggggtctgctcaggaacagtgaag 910
DB 301 TGAAGTCCCCCGTGGACTACGGCAGCAGTCCCGGGGTCTGGTCTCAGGACAAAGTGAAG 360

QY 911 ggaagtttgatgtccagtggaattttgttaagatgtaccctaaccagctccgcacaca 970
DB 361 GGAAGTTTGATGTCCAGTGGATTTTGTAAAGATGTACCAATACCAGTCCGGCACA 420

QY 971 tcaggtcgagataacgacaaacacccggtcacaactcccgagacccagcaggtgc 1030
DB 421 TCAGGCTGGAGATAACGACAAACAAACCGGTCAACAACTCCCGGGACACCCAGGAGTGC 480

QY 1031 ccttgaataaagcgaagctgctgaataattatcagttccctacaagcacacaactcca 1090
DB 481 CTTTGAATAAAGCAAGCAAGTCTGTAATAATTATCAGTTCTCTACAAGCACAAAGTCCA 540

QY 1091 tcttcagcacttctgctcactacgagaagcgcca-gaggaggagggtggtgcgaag 1149
DB 541 TCTTCAGCACTTTGCTCACTACGAGAGCGCCAGGAGGAGGAGGTGTCGCGCAAG 600

QY 1150 aacggcagagtcgaacaaa-caatgaggcggaaccagtttcttacctatcttaacgctt 1208
DB 601 AACGGCAGAGTCGAACAAACCAATGAGGCGCAACCAAGTATCTTACATGTTCTTAAGCTTT 660

QY 1209 gacttttga--aaacagttttaaaccagtgctgtgtcagctccagtggtcgtccgtg 1266
DB 661 GACTTTGAACAACAGCTTTAAACCACTGTCGCTTGGTCAGTCCAGTGTGTCGTCGCCGTG 720

QY 1267 cgggggtgagtggtgcatctcttgccttcttctgtcttg---attttggccagatgga 1322
DB 721 CGGGGTTGAGTGTGCTCTTTGGCCCTTTCATGTCGGGTGAATTTTGGCCCGCATGGG 780

QY 1323 tctgcatttatctgactcttcttctatgatatataacacctgt 1363
DB 781 TCTGCCCTTCATTGGAGCTTCTCCACAGGCATATAACCCGT 821

RESULT 2

BE92982
LOCUS 601434159F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919118 5',
DEFINITION mRNA sequence.
ACCESSION BE92982
VERSION BE92982.1 GI:10353885
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 852)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCPD/BTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Place: LHAM5748 row: f column: 15
High quality sequence stop: 643.

FEATURES

source
1..852
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3919118"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 223 a 199 c 167 g 263 t
ORIGIN

Query Match 32.1%; Score 650.8; DB 10; Length 852;
Best Local Similarity 94.9%; Pred. No. 2.2e-119;
Matches 748; Conservative 0; Mismatches 32; Indels 8; Gaps 7;

QY 969 catcagcctggagataacgacacaaacccggtcacaaactccggagacaccagaggt 1028
DB 1 CATCAGCTGGCAANTAACGACAAACACCGGTCAACAACTCCCGGACACCCAGGAGGT 60

QY 1029 gcccttagaaaaagccaagcaagtgctgaaaaattatcagttcctacaagcacacaacctc 1088
DB 61 GCCCTTAGAAAAAGCCAAAGCAAGTGTGAAAAATTATCAGTTCTTACAAGCACACAACCTC 120

QY 1089 catcttcgacgaacttgcctcactacgagaagcgcca-gaggagagaggtggtgcgcaa 1147
DB 121 CATCTTCGAGCACTTTGCTCCTACGAGAAGCGCCAGGAGGAGGTGGTGGCGCAA 180

QY 1148 ggaacggcagagtcgaaacaaacaaatgagggcgaaaccagttcttacctatcttaacgtt 1207
DB 181 GGAACGGCAGAGTCGAAACAAACAAANTGAGGCGCAACCAAGTTCTTACATGTTCTAAGGTT 240

QY 1208 tgactttgaaaaacagttttaaacaacgctgtcttggttcagctccagtgctgcctccgtgc 1267
DB 241 TGACTTTGAAAAACAGTTTAAACACAGCTGCTTTGGTCAGCTCCAGTGTGTCGCCGTGC 300

QY 1268 ggggggtgagtgctgcatcttctgccttcttgcgttgatttttggccagatggatctgc 1327
DB 301 GGGGGTTGAGTGTGCAATCTTTCCTTTCGTTGATTTTGGCCAGATGGATCTGC 360

QY 1328 atttattgtacttttctatgtattataatcctgtagaagtcacataaagaagagatt 1387
DB 361 ATTTATTGTACTTTTCTATGTATTATATCTCTGTAGAGTCACTAATAAAGAGTA-T 419

QY 1388 ttttttgcagcttatcaatcagactgatctaatgtgaagttaagtatcccttaaaac 1447
DB 420 TTTTGTGTGTCAGCTTATCAATCAGACTGATCTAATGTGAAATGTAAAGTATCTCTTAAAAAC 479

QY 1448 aaagcatctattttggcagaaaattgtgttcttaaatcagctcatttgatat-tctgtgag 1506
DB 480 AAAGCATCTATTATTTGGCAGAAATTTGTTCTTAAATTCAGTCAATTTGATATCTCTGTGAG 539

QY 1507 acttcataattctcatcctttattgcttttttagcaaacataaagaaccatgagtcattt 1566
 |||||
 Db 540 ACTTCATATTTCTCATCCCTTATTGCTTTTAGCAACACATAAGAACCATGAGTCATT 599
 |||||
 QY 1567 tgcatttagagtgattctcgataaatactctctgaaataactgaaatacaaaaggttaagt 1626
 |||||
 Db 600 TGTCAATTTAGAGTAGTCTCGATAAAATCTCTTGAATAACTGAAATCAAA--GGTTAATGA 657
 |||||
 QY 1627 ttttggcttcattgattgctcattttatttattctgttatcggtctaa-agtgcctaattta 1685
 |||||
 Db 658 TTCTTGTCATTCATGATTTGCTCTTATCATCTGCAATCGGGCTTAACAGTGCTACTTTA 717
 |||||
 QY 1686 -ccatttgatttttctgtctagacagataaacttttaattttcaatttggcagacatt 1744
 |||||
 Db 718 CCCCATTCGAGCTCTCTGCTAGACCGCTTACCCCTT-ATTTTCCACATTCGGCCGACCTTT 776
 |||||
 QY 1745 tttttttt 1752
 |||||
 Db 777 TCTTCTCT 784
 |||||

RESULT 3

AV079675
 LOCUS AV079675 1014 bp mRNA linear EST 24-OCT-2001
 DEFINITION AV079675 Mus musculus stomach C57BL/6J adult Mus musculus cDNA
 clone 2210410K23, mRNA sequence.

ACCESSION AV079675
 VERSION AV079675.2 GI:16381140
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1014)
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)

TITLE On Jun 25, 1999 this sequence version replaced gi:5211123.
 JOURNAL Contact: Yoshihide Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.
 Computational Analysis of Full-Length Mouse CDNA Sequences Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES

Location/Qualifiers
 1..1014
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2210410K23"
 /clone_lib="Mus musculus stomach C57BL/6J adult"
 /sex="male"
 /tissue_type="stomach"
 /dev_stage="adult"

BASE COUNT 266 a 238 c 252 g 256 t 2 others
 ORIGIN

Query Match 32.0%; Score 648.8; DB 9; Length 1014;
 Best Local Similarity 82.2%; Pred. No. 5,4e-119;
 Matches 822; Conservative 0; Mismatches 159; Indels 19; Gaps 6;

QY 406 cctctccagacagcccccagcttggctcaaccgcagctatcagagccctcagagcca 465
 |||||
 Db 19 CCTCTCCCTGTNCAGCCCCNACCTGTGTGTC-CAGCCACAGTATCAGAGCCCTCAGCAGCCA 77
 |||||
 QY 466 cccagacccctgggtgccccacgacacgaacagcgcttggcgagagcgagg 525
 |||||
 Db 78 CTTCAACCCCGCTGT-TGGCTCTCTGAAACAGAAATGACAGATGTGCGCAGAGTGGAGGG 136
 |||||
 QY 526 gctggcagcgatagcaactctctggaaacgtccagcctaatctgccccagcgtcgaa 585
 |||||
 Db 137 GCCAACAGTGCAGTAACCTCTGTTGGAAATGCCAACCTTACTTCTGCCCCGAGTGTAGAA 196
 |||||
 QY 586 tccaccccgctcttgaaaaactggaagcgctgcacagctacacacccgaaagagttgag 645
 |||||
 Db 197 TCCCAACCTGTCTGTGAGAAACTGAAAGCTGCCACACAGCTATAACCTTAAGAGTTCGAC 256
 |||||
 QY 646 tggaaatgaaacgcgcgctgttctcatcatcaagagctactctgagacacatccac 705
 |||||
 Db 257 TGGAAATCTTAAGAGTGGGGGGGTGTTCATCATCAGAGAGCTATTCAGAGACCATCCAC 316
 |||||
 QY 706 cgtccatttaagtactccatctggtgtgtagcacagagcagcgcaacaagcgctggacagc 765
 |||||
 Db 317 CGTCCATCAAGTACTCCATCTGTTGTAGTAGTGAACACGGCAACAAGCGCTGGACGCG 376
 |||||
 QY 766 gcttcgcgctgcatgagcagaagggggccgctctaccctgctcttcagcgtcaatggagct 825
 |||||
 Db 377 GCCTTCCTCCCTCATCAGCAGCAAGGGGCGCTGTTTATCTCTCTTCAGTGTCAATGGAGT 436
 |||||
 QY 826 gggcatttttgggggtggcgagatgaagtcctcccgctgactacgacacagtcgaggg 885
 |||||
 Db 437 GGACATTTCTGTGGGGTGGCAGAGATGAAGTCCCTCTGTGGACTATGGCACCACCGCTGGG 496
 |||||
 QY 886 gtctggtctcaggacaagtgggaagggaagtttgatgctcagtgagatttttttaaggat 945
 |||||
 Db 497 GTCTGGTCTCAGSACAAGTGGGAAGGAAAAGTTTGATGTGAAGTGGATTTTGTGAAGGAT 556
 |||||
 QY 946 gtaccataaacagctccggcacatcaggtctggagaataacagcaacaaacccggtcaca 1005
 |||||
 Db 557 GTGCCCAACAACACCTCGGCACATCAGACTGGGAATTAACGACAAACAACTGTGCACA 616
 |||||
 QY 1006 aactccccggacacccagaggtgcctctagaaaaagccaagcaagtcgtgaaattatc 1065
 |||||
 Db 617 AACTCCCGTGATACACAGGAGGTGCCCTTAGAAAAAGAAAACAACTGCTGAAGATTATC 676
 |||||
 QY 1066 agttcctacaagcacacaacacctccalcttccgacgactttgtcactacgagaagcgcca- 1124

```

Db 677 GCTTCCTATAAGCACACACCTCTATCTTTGACGACTTTTCATTATGAGAAGCCCGAG 736
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1125 gaggagagagagtggtgcgcaagcagcgagcgcgaaacaaacaaatgagggcgaaac 1184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 GAGGAAGAGAGGTGTCGTAAAGAAACACACAGATCGAAACAAACAAATAGAACAGCC 796
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1185 agtttctacatggtttaaactgttgactttgaaacagttttaaacacacgtgtgcttggtc 1244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 AGTTTCTTTTGGTTAATGGTTGACTTTTGAACACAGATTTTAAAGCTGTATGCTTGGTG 856
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1245 agct-----ccagtgtgtcccggtgcggtggttgaagtggttgcatttgcct 1293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 CTGTCTCCGAGTCAGTCACAGTGTCTCTCGTGCCTGCGGGTTCATTTGCATCTTTTATCT 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1294 ttcttgcgttgattttccagatggtatctgcatatttttgcatttttctctatctatt 1353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 917 TTGTAG----TTCAATTTTTCAGATGATCGATTCATTTGTGA-TTTTTCATATGATTT 971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1354 ataactctagaagtcactaataaaggagttattttttt 1393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 972 ATAATATTCTAGAACTCACTAATAAAGGAGTATTTTGTGTTT 1011
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
AL574115/c 661 bp mRNA linear EST 16-FEB-2001
LOCUS AL574115 LTI_NFL006_PL2 Homo sapiens cDNA clone CSDDI040YK06 3
DEFINITION prime, mRNA sequence.
ACCESSION AL574115
VERSION AL574115.1 GI:12934008
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 661)
AUTHORS Li,W.B., Gruber,C., Jesse,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDDI040YK06"
/tissue_lib="LTI_NFL006_PL2"
/feature="vector: pCMVSPORT 6; Site 1; NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 119 a 179 c 178 g 181 t 4 others
ORIGIN

Query Match 31.8%; Score 644.6; DB 9; Length 661;
Best Local Similarity 99.1%; Pred. No. 3.7e-118;
Matches 655; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy 527 ctggcagcagcagcactctcctgaaacgctccagcctaattctgccccagcgtcgaaat 586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CTGGCAGCATAGCAATCTCTCTGGAAGCTCCAGCCTAATTTCTGCCCCAGCGTGAAT 602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Qy 587 ccacccccctcttgaaactgaagcgtgctcacagctacaaacccgaaagagtttgagt 646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CCCACCCCGCTCTTGAAGAACTGAAGGCTGCTCACAGCTACAACCCGAAAGAGCTTTGAGT 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 647 ggaatctgaaagcggcggtgtgttctatcatcaagagctactctgaggagcagatccacc 706
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GGAATCTGAAAGCGGGCGGTGTCTCATCATCAAGAGCTACTCTGAGGAGAGATCCACC 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 707 gctccatttaagtaactcctctgtgtgtagcacagacacgcaacaaagcgctgtgagcagcg 766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 GCTCCATTAAAGTACTCCATCTGCTGTAGCACAGACAGCGCAACAACGCGCTGGACAGCG 422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 767 ccttcgcgtgcatgagcagcaggggcccgtctacgtctcttcagcgtcaatgggagtg 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CTTTCGCTGCATGACGACGACAGGGCGCGTCTACCTGCTCTNCAGCGTCAATGGGAGTG 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 827 ggcatttttgggtggcgagagatgaagtcctccctgagactacgacacagtcgcccgggg 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GGCATTTTTCGCGGTGGCGGAGATGAAGTCCCGCTGGACTACGGCACCGAGTGCCTGGG 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 887 tctggtctcagacaaagtgaaggggaagtttgatgtccagtgatgtttttttaaagatg 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TCTGCTCTCAGAACAAAGTGAAGGGGAAGTTTGATGTCCAGTGGATTTTGTGTAAGATG 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 947 tacccaataaccagctccggcacatcagcgtgagagataacacgacacaaacccggtcacaa 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TACCCAATAACCAAGCTCCGGCACATCAGGCTGGAGATAAGAGACAACAAACCGGTCAACA 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1007 actcccgagacacccaggggtgccccttagaaaaaaccaagcgaagtgtgaaatattca 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ACTCCCGGACACCCAGAGGTGCCCTTAGAAAAAGCCAAAGCAAGTCTCTGAAATTTATCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1067 gttcctacaagcacacaactcctctctgagcactttgctcactcagagagcgcaa-g 1125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GTNCTACAAAGCACACAACCTCCATCTTCGACACATTTGCTCACTACGAGAACGCCAGG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1126 aggagagagaggtgtgtgcgcaaggaacgagcagagtcgaacaaacaaatgaggcgcaacca 1185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AGGAGGAGGAGGTGTGTGCCAAGGAACRCAGAGTCGAAACAAACAATGAGGCGCAACCA 2
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1186 g 1186
Db 1 g 1

RESULT 5
AL515908/c 1068 bp mRNA linear EST 13-FEB-2001
LOCUS AL515908 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA001YB18 3
DEFINITION prime, mRNA sequence.
ACCESSION AL515908
VERSION AL515908.1 GI:12779401
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1068)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..1068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DA001YB18"
/tissue_lib="LTI_NFL011_NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"

FEATURES
source

```



```

Db 323 CCTCCATCTTTTGACGACTTTGCTCCTACTAGGAGAAAGCGCCAGGAGGAGGAGGTGGTG 264
Qy 1143 cgcaggaacgagcagagtcgaacaaacaaatgagggcggaacacagttcttacatgttcta 1202
Db 263 GCACAGGAACGGCAGAGTCGAACAAACAATGAGGCGGAACAGATTCTTACATGTTCTA 204
Qy 1203 acgtttgactttgaaacacagtttaaaacacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1262
Db 203 ACCTTTGACTTTGAAACACAGTTTAAACACAGCTGTGCTGTGCTGCTGCTGCTGCTGCTC 144
Qy 1263 cgtgagggaggttgaggttgagtcatttgccttcttcttcttcttcttcttcttcttctt 1322
Db 143 CGTGGCGGGGTTGAGTGTGTCATCTTTTGGCTTTTCTTTGCTGCTGCTGCTGCTGCTGCT 84
Qy 1323 tctgcatttatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1382
Db 83 TCTGCATTTATTTGTAATTTTCTATGTATTAATTAATCCCTGTAGAACTCACTAATAAGGA 24
Qy 1383 gtatttttttttgcagcttatca 1406
Db 23 GTA-TTTTGTGTCAGCTTATCA 1

RESULT 7
AV726565 932 bp mRNA linear EST 17-OCT-2000
LOCUS AV726565 HTC Homo sapiens cDNA clone HTCAxE08 5', mRNA sequence.
ACCESSION AV726565
VERSION AV726565.1 GI:10835986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
Gu, X., Peng, Y., Song, H., Huang, Q., Yang, X., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: zeguang han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. .932
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCAxE08"
/clone_lib="Htc"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 240 a 166 c 179 g 340 t 7 others
ORIGIN
Query Match 30.9%; Score 626.6; DB 9; Length 932;
Best Local Similarity 97.0%; Pred. No. 1.4e-114;
Matches 679; Conservative 0; Mismatches 17; Indels 4; Gaps 4;
Qy 1329 ttatttgcagcttatcaatcagactgatctaagtgtgaatgtgaatgtgaatgtgaatgtgaatgt 1388
Db 2 TTTACTCGTACTTTNCTATGATTATTAATCCTGTAGAGTCACTAATAAGGAGTA-TT 60
Qy 1389 tttttgtcagcttatcaatcagactgatctaagtgtgaatgtgaatgtgaatgtgaatgtgaatgt 1448

```

```

Db 61 TTTTGTGTCAGCTTATCAATCAGACTGATCAATGTGAATGTAAAGTATCTTTAAACA 120
Qy 1449 aagcalctatttggcagaaattgttcttaaatcaagtcatttgaatattctgtgagac 1508
Db 121 AACATCTATTTTGGCAGAAATTTGTCTTAAATTCAGTCATTTGATATCTGTGAGAC 180
Qy 1509 ttcataattctcatccctttattgcttttttagcaacaataagaacacatgagtcattttg 1568
Db 181 TTCATATTTCTCATCCCTTTATTTGCTTTTAGCAACATAAGAAACCATCAGTCATTTTG 240
Qy 1569 tcatttagagtaacctgataaaatctcttgaaataactgaaatcactgaaatcactgaaatcactg 1628
Db 241 TCATTTAGAGTATTCGTATAAAATCTCTTGAAATACTGAAATCAAAAGCTTAATGATTT 300
Qy 1629 ttgttcattctctgatttgcattttatttattcttcttcttcttcttcttcttcttcttctt 1688
Db 301 TTTGTTCATTTCTGATTGTGTCATTTTATCTCTTATCTGCTTAAAGTCTGCTAAATTTTACC 360
Qy 1689 atttgattttctctgtagacagataaacttttaatttttcaaaatttggcagacactttttt 1748
Db 361 ATTTGATTTTCTGCTAGACAGATAACTCTTAATTTTCAATTTTGCACACACTTTNTT 420
Qy 1749 ttttttttgaaatcttcttcttccagatctgttgcacactgacagccaccgcctccctc 1808
Db 421 TTTTGTGCTTAAATCTTTCTTCCAGATCTGTGTCACACTGAAACAGCCACCCGTCCTC 480
Qy 1809 actgtcctgggtccgcatgggctggatggatggatggatggatggatggatggatggatggat 1868
Db 481 ACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Qy 1869 agtgctctttagtctgtgttcagggctgggcatctcttcttcttcttcttcttcttcttctt 1928
Db 541 ACCTGCTTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
Qy 1929 ttacacaccttttcttaagaattcctaagcgtcttcttcttcttcttcttcttcttcttctt 1988
Db 599 TTTTACACCTTTTCTAANGACATCTAATGCGCTCTTAAG-TTTTATACCAATAATGCTGAG 657
Qy 1989 ctttaagttagagctgtgtgtagtacagacagtgatgaga 2028
Db 658 CTTTAAAGTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 697

RESULT 8
BM423045 1479 bp mRNA linear EST 29-JAN-2002
LOCUS PLATE1_D11 Rhesus Macaca mulatta cDNA, mRNA sequence.
DEFINITION PLATE1_D11
ACCESSION BM423045
VERSION BM423045.1 GI:18392539
KEYWORDS EST.
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
REFERENCE 1 (bases 1 to 1479)
AUTHORS Katze, M.G., Bumgarner, R., Korth, M., Feldman, R., Amjadi, M. and
Holzman, T.
Expressed sequence tags from Rhesus macaque spleen
Unpublished (2002)
Contact: Holzman T
Katze Lab
University of Washington
Box 358070, Seattle, WA 98195-8070, USA
Tel: 206 732 6156
Fax: 206 732 6055
Email: ted@locke.hs.washington.edu
PLATE1_D11 1479 bases; 23.2 mean phred score.
FEATURES
Location/Qualifiers
1..1479
Source
/organism="Macaca mulatta"
/db_xref="taxon:9544"

```

AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .									
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),									
JOURNAL	Tumor Gene Index									
COMMENT	Unpublished (1997)									
	Contact: Robert Strausberg, Ph.D.									
	Email: cgaps-r@mail.nih.gov									
	Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,									
	M.D., Louis M. Staudt, M.D., Ph.D.									
	cDNA Library Preparation: M. Bento Soares, Ph.D.									
	cDNA Library Arrayed by: Greg Lennon, Ph.D.									
	DNA Sequencing by: Washington University Genome Sequencing Center									
	Clone distribution: NCI-CGAP clone distribution information can be									
	found through the I.M.A.G.E. Consortium/LLNL, send email to:									
	info@image.llnl.gov									
	Seq primer: -40UP from Gibco									
	High quality sequence stop: 459.									
FEATURES	Location/Qualifiers									
source	1. 664									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/clone="IMAGE:3292137"									
	/clone_lib="NCI_CGAP_C111"									
	/tissue_type="B-cell, chronic lymphocytic leukemia"									
	/lab_host="DH10B"									
	/note="Vector: pT73D-Pac (Pharmacia) with a modified									
	polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA									
	was primed with a Not I - oligo(dT) primer [5,									
	TGTTACCAATCTGAAGTGGAGCGCGCATGTGTTTTTTTTTTTTTTTTTTTTTTTTT									
	T 3']; double-stranded cDNA was ligated to Eco RI									
	adaptors (Pharmacia), digested with Not I and cloned into									
	the Not I and Eco RI sites of the modified pT73 vector.									
	Library is normalized, and was constructed by Bento									
	Soares and M.Fatima Bonaldo."									
BASE COUNT	168 a	177 c	150 g	168 t	1 others					
ORIGIN										
	Query Match	29.7%;	Score 601.8;	DB 10;	Length 664;					
	Best Local Similarity	98.0%;	Pred. No. 1.2e-109;							
	Matches 641;	Conservative	0;	Mismatches 8;	Indels 5;	Gaps 3;				
Qy	745	ggcaacaagcgcctggacagcgcttcgcgctgcatgagcagcaa	---	ggggcccgctctac	801					
Db	664	GGCAACAAGCGCTGGACAGCGCTTCGCTGCATGAAGCAGCAAGGGCCCGCTCTAC	605							
Qy	802	ctgctcttcagcgtaaatggagtgggcaatttttggggtggccgagatgaagtcctccc	861							
Db	604	CTGCTCTTCAGCGTCAATGGGAGTGGGCATTTTGTGGGTGGCGGAGATGAAGTCCCCC	545							
Qy	862	gtgactacgacccagtcgcggggtctggtctcagggacaagtgaagggaagtgtgat	921							
Db	544	GTGNACTACGGCACAGTCCCGGGGTCTGGTCTCAGGACAAAGTGAAGGGGAAGTTGAT	485							
Qy	922	gtccagtggaattttgttaagatgtacccaataaccagctccggcacatcaggtggag	981							
Db	484	GTCCAGTGGATTTTGTAAAGGATGTACCAATAACCAAGTCCGCGACATCAGGCTGGAG	425							
Qy	982	aataacgacaacaaccgggtcacaaactccggggacacccaggaggtgcccttagaaaaa	1041							
Db	424	AATAACGACAACAACCGTCAACAAACTCCCGGGACACCCAGGAGGTGCCCTTAGAAAAA	365							
Qy	1042	gccaagcagtgctgaaatctatcagttcctacaagcacacaacctccatcttcgacgac	1101							
Db	364	GCCAAGCAAGTGTCTGAAATTAATCATGTTCTTACAAGCACACANACCTTCCATCTTCGACGAC	305							
Qy	1102	tttgctcactacgagaagcgcca-gaggagaggaggtgtgtgcgaaggaacggcagagt	1160							
Db	304	TTTGCTCACTACGAGAAGCGCCAGGAGGAGGAGGTGGTCCGCAAGGAACGGCAGAGT	245							
Qy	1161	cgaaacaaacaaatgaggcgcaaccaggttcttcatgttctaaacgttttgacaaa	1219							
Db	244	CGAAACAAACAATGAGGGCGAACCAAGTTCTTTTACATGTTCTTAACGTTTGACTTTGAAAA	185							

Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 124 a 169 c 151 g 156 t
ORIGIN

```
Query Match      28.8%; Score 583.2; DB 9; Length 600;
Best Local Similarity 99.3%; Pred. No. 6.1e-106;
Matches 596; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 676 atcaagagctactctgagcagcatccacccgctccatcaagtaactcctatctggtgtagc 735
Db 600 ATCAAGAGCTACTCTGAGGAGCATCCACCGCTCCATTAAAGTACTCTCCATCTGGTGAGC 541

Qy 736 acagagcagcgcaacagcgctctgagcagcgctctccgctcgtatgacgacgaagggccc 795
Db 540 ACAGAGCAGCGCATCAAGCGCTGGACAGCGCTTCCGCTGCATGAGCAGCAAGGGGCC 481

Qy 796 gtctacctgctctcagcgtcactgagtgagtgagtgatctttttggtgggtggccgagatgaag 855
Db 480 GTCTACCTGCTCTTACGCGTCAATGGGAGTGGGCATTTTCTGGGGTGGCCGAGATGAAG 421

Qy 856 tccccggtggactcggccacagtcgcggggtctggtctcagcagcagtggaaggggaag 915
Db 420 TCCCCGCTGGACTACGGCACCAGTGCCTGGGGTCTGGTCTCAGGACAACTGGAAGGGGAG 361

Qy 916 ttgatgtccagtggatttttgaagatgtacccaataaccagctccgcgcacatcagg 975
Db 360 TTTGATGTCCAGTGGATTTTGTAAAGGATGTACCCAATAACCAGCTCCCGGTACATCAGG 301

Qy 976 ctgagataacgacaacaccgggtcacaaactcccggggacacccagagaggtgccctta 1035
Db 300 CTGAGATAAACGACAACAAACCGGTTCACAAACTCCCGGGACACCCAGGAGTGCCTTTA 241

Qy 1036 gaaaagcgaagcagtgctgaaatctatcagttctcacaagcacaacacctccatcttc 1095
Db 240 GAAAAGCCAAGCAAGTGTCTGAAAATATATCAGTTTCTTACAGCAGCACAACTTCCATCTTC 181

Qy 1096 gacgactttgctactcagagaagcgcca-gaggagagggaggtggtgcgcaagaagcgg 1154
Db 180 GACGACTTTGTCTACTACGAGAGCGCCAGGAGGAGGAGGTGGTGGCGCAAGCAAGCG 121

Qy 1155 cagagtcgaacaacaatgagggcgcaaccagtttcttaacatgttctaactttgacttt 1214
Db 120 CAGAGTCGAACAACAATGAGGGCGAACCAGTTTCTTACATGTTCTTAACGTTTGACTTT 61

Qy 1215 gaaaacagtttaaaacacgtgctgtggtcagctccagtgctgctcccgctgcgggggtt 1274
Db 60 GAAAACAGTTTAAACACGTGTGCTGTGCTCAGCTCCAGTGTGCTGCCGTGCCGGGGTT 1
```

```
RESULT 12
BE892829 601433961F1 NIH_MGC_72 Homo sapiens CDNA clone IMAGE:3919191 5',
DEFINITION mRNA sequence.
ACCSSION BE892829
VERSION BE892829.1 GI:10353393
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 808)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
```

Tissue Procurement: ATCC/DCTD/DFP
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9748 row: i column: 16
High quality sequence stop: 648.

```
FEATURES
source
Location/Qualifiers
1..808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3919191"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DHI0B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SpORF6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 222 a 172 c 163 g 251 t
ORIGIN
```

```
Query Match      28.7%; Score 581.8; DB 10; Length 808;
Best Local Similarity 88.8%; Pred. No. 1.2e-105;
Matches 711; Conservative 0; Mismatches 77; Indels 13; Gaps 7;

Qy 977 tggagaataacgacacaacacccgggtcacaaactccgggacacccagagaggtgccttag 1036
Db 10 TGGAGAATAACGACACAACAAACCGGTTCACAACTCCCGGACACCCAGGAGGTGCCTTAG 69

Qy 1037 aaaaagcgaagcgaagtgctgaaattatcagttctcacaagcacaacacctccatcttcg 1096
Db 70 AAAAGCCGAAGCAAGTGTCTGAAAATTTATCAGTTCTTACAGCAGCACAACTCCATCTTCG 129

Qy 1097 acgactttgctcactacgagagcgcca-gaggagagagaggtggtgcgcaaggaacgac 1155
Db 130 ACGACTTTGCTCCTACGAGAGCGCCAGGAGGAGGAGGTGGTGGCAAGGAGCGC 189

Qy 1156 agagtcgaacaacaactaggggggaacccagtttcttaacatgttctaactttgactttg 1215
Db 190 AGAGTCGAACAACAATCAGGGCGCAACAGTTTCTTACATGTTCTTAACGTTTGACTTTG 249

Qy 1216 aaaaacagtttaaaacacagtgctgttggtcagctcagtgctgcctccgctgcgggggtt 1275
Db 250 AAAACAGTTTAAACACAGTGTGCTGGTCCAGCTCCAGTGTGTCGTCGCGGGGGGTG 309

Qy 1276 agtgtgcactctttgctcttctgtctgtgattttggccagatggatctgcatattatt 1335
Db 310 AGTGTGTCATCTTTGCCCTTCTTGTGCTGTGATTTTGGCCAGATGGATCTGCATTTATTT 369

Qy 1336 gtaacttttctatgtattataatcctctgtagaagtcactaataaaggagattttttttg 1395
Db 370 GTACTTTTCTATGTATTATAATCTGTAGAGTCCTTAATAAGGAGTA-TTTTTTTTGG 428

Qy 1396 tcagcttatcaatcagactgacttaagtgaataagatccttaaaaaaacaagaacatc 1455
Db 429 TCAGCTTATCAATCAGACTGATCTAATGTGAATTAAGTATCTTAAACCAAGCAATC 488

Qy 1456 tatttggcagaaaattgttctttaaattcagctcatttgatattctgtgagacttcatat 1515
Db 489 TATTTGGCAGAAATTTGTTCTTAATTCAGTCATTTGATATTCTGTGAGACTTCATA- 547

Qy 1516 ttctaatccctttattgttttttagcaacaataagaagaacctaga-----gtcattttgtc 1570
Db 548 TTCTCATCCCTTTATTGCTTTTGTAGCAAAACATAAAGAAAACCAATGAGCTCATTTGGTC 607

Qy 1571 atttagagattctctgataa-aatctctgaaaataactgaaatacaaaaggttaaatgatttt 1629
Db 608 ATTTAGAGTATTCTGTATAACAATCTCTTGAAATACTGGAACCTCCACAGGGTTATGATTTT 667
```

```
QY 1630 ttgt-tcaattgattgttcattttatttattctgttctatcggtctaaagtgctaatccaacc 1688
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 CCGTGCATCTGATTGGCCCTTAAATCTGTATCGGGCCAAAAGTCTATTCTCC 727

QY 1689 atttgattttctgctagacagataacttttaatttttcaaaatttggcagacactttttt 1748
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ATTGG---TTTCTGGGTGACAGATAACTTTTACTTCTCCAATTGGCGCACACCCTTTGT 784

QY 1749 ttttttttgaaaatttttcc 1769
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 GGAACTCTCTCAATTGTGCC 805

RESULT 13
BI091481
LOCUS 60285860F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5000127 5',
DEFINITION mRNA sequence.
ACCESSION BI091481
VERSION BI091481.1 GI:14509811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs.r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11031 row: h column: 16
High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source           1..736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5000127"
                     /clone_lib="NIH_MGC_10"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /note="organ: cervix; vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5 Kb. Library prepared by Life
                     Technologies."
BASE COUNT          193 a 123 c 146 g 274 t
ORIGIN
Query Match          28.6%; Score 580.4; DB 10; Length 736;
Best Local Similarity 96.4%; Pred. No. 2.2e-105;
Matches 690; Conservative 0; Mismatches 16; Indels 10; Gaps 9;

QY 1313 ccagatgagctgacttatttattgtactttttctatgtattataatccctgtagaagcacc 1372
Db 1 CCAGATGGATCTGCATTTATTTGTACTTTTCTATGTATTATAATCCTGTAGAGTCAC 60

QY 1373 taataagagatttttttctgctatcaatcacactgactaatgtgaaatgta 1432
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TAATAAGGAGTA--TTTTTTGTGAGCTATCAATCAGACTGATCTAATGTGAAATGTA 118

QY 1433 agtaccttaaaacaaagcatctattttggcgagaattgtgtctttaaattcagtcatt 1492
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 AGTATCCTTAAAAACAACAGCATCTA--TTTGGCAGAAATGTGTCTTTAAATTACAGTCATT 177

*QY 1493 tgatatctgtgagacttcatatcttctcccttttctgtcttttttagcaaacataagaa 1552
```

```
Db 178 TGATATCTGTGAGACTTCATATTTCATCCCTTTATTGCTTTTATAGCAACATAAGAA 237
QY 1553 accatgagtcatttctcatttagagattcttgataaaaaatcctctgaaaaactgaaatc 1612
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 ACCATGAGTCATTTGGTCATTTAGAGTATTCTGATAAAATCTCTTGAANAATCTGAAATC 297
QY 1613 aaaaagttaatgattttttgttcattctgatttgcatttttattatctgtatcggctcta 1672
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 AAAAGCTTAATGA-TTTTGTTCATTCGATTGTGTCATTTTATTATCTGTTATCGGCTCA 356
QY 1673 aagtgttaattaccatttgattttctgtctagacagataaacttttaatttttccaaatt 1732
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 AAGTGTCAATTTACCATTTGATTTTCTCTGTAGACAGATAACTTTTAATTTTCAAAATT 416
QY 1733 tggcagacacttttttttttttttttttttttttttttttttttttttttttttttttt 1792
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 TGCACAGACAC-TTTTGTGAAAATCTTCCTCTCCAGATCTGTGGCCCACTGAA 475
QY 1793 cagccacccgtccctcactgtcctggtgccgattgggtggatgggtggtgggcatgat 1852
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 CAGCCACCCGTCCTCACGTCTCTGCTGCTCCGATGGGCTGGATGCTGTGGGGCATGAT 535
QY 1853 gttgagagaactgggaagtgcttttgggtcaggttcaggttcaggttcaggttcaggttc 1912
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 GTGTGAGGAGAACT-GAAGGTGCTTAAAGTCTGTGTCCA-GGTCTGGGCAATTC-TTGGTGGTT 592
QY 1913 gcaactcttttaaaattttacacctttttttaaagaattccttaagtcgtcttaagttttta 1972
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 GCATCT-TTTTAAATTTGACACCTTTTCTTTAAAAATCTTAATAGCCCG-CTGAAGTTTTTA 650
QY 1973 taccataatgctgagcttttaagttaggactggttagtacagacagtgtagtggga 2028
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 TACCAATAATGCTGACGCTTAAGTGTAGGATCTGGTAGTACACACACTGTGTGATGGA 706

RESULT 14
AW964022
LOCUS EST376095 MAGE resequences, MACH Homo sapiens cDNA, mRNA sequence.
DEFINITION AW964022
ACCESSION AW964022
VERSION AW964022.1 GI:8153858
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 669)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnqu@igr.org
Plate: 196
Seq primer: Reverse.
Location/Qualifiers
     source           1..669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MACH"
                     /note="Vector: pBluescriptSkm"
BASE COUNT          175 a 107 c 135 g 252 t
ORIGIN
Query Match          28.2%; Score 571; DB 9; Length 669;
```

Best Local Similarity 97.4%; Pred. No. 1.6e-103;		Matches 591; Conservative 0; Mismatches 15; Indels 1; Gaps 1;	
Qy	1396	tcagcttatacaatcagactgatcctaagtgaatgtaagtatccttaaaacaagcatic	1455
Db	1	TCACGTATCAATCAGACTGATCTAATGTGAATGTAAGTATCCTTAAAAACAAGGATC	60
Qy	1456	tattttggcagaaattgttctttaaattcaagtcatttgatattctctggagactcatat	1515
Db	61	TATTTTGGCAGAAATTGTCTTAAATTCAGTCATTGTGATATCTGTGAGACTTCATAT	120
Qy	1516	tctcatccctttattgtcttttttagcaaaacataagaacaacatgagtcattttgtcattta	1575
Db	121	TTTCATACCCCTTTATGCTTTTATAGCAACATAAGAACCATGATGATCTTTTGTCAATTA	180
Qy	1576	gagtatctgataaaatctcttgaaatactgaaatacaaaagggttaagtatttttttttc	1635
Db	181	GAGTATCTGTGATAAATCTCTGAAATACGAAATCAAAAGGTTAATGATTTTTTGTTC	240
Qy	1636	attctgatttgcattttattattctgttctggtctaaagtgtctaattttaccatttgc	1695
Db	241	ATTCTGATTTGTCATTTTATTTATCTGTTATCGGTCTAAAGTGCTAATTTACCCATTTGAT	300
Qy	1696	ttttctgtgagacagataactttttaatttttcaaattttgcagacac-ttttttttttt	1754
Db	301	TTTTTCTGTAGACAGATAACTTTTAAATTTTCAAAATTTGGCAGACACTTTTTTTTTTTT	360
Qy	1755	tttgaaatcttcttccagatctgttcccactgaacagcaaccgctccctcaactgc	1814
Db	361	TTTGAAATCTTCTCCTCCAGATCTGTGGCCACTGAACGCCACCCGCTCCCTCACGTGC	420
Qy	1815	ctgggtccgattgggtggatggtgttggggcatgagtgtgaggaactggaaggtgc	1874
Db	421	CTGTGTCCGATTTGGCTGGATGCTGTGGGCATGATGTGTGAGGAACCTGGAAGTGC	480
Qy	1875	tttaggtctggttcagggtcgggcatctctttgttggttgacatctcttttaaattttaca	1934
Db	481	TTTAGGCTGTGGTTCCAGGTCGGGCATACTTTGTGCTGTGGCACATCTTTTAAATTTTACA	540
Qy	1935	cttttcttaagaattctaagtcgcgtctaaagtcttttataccaataatgctgagctttaa	1994
Db	541	CCGTGTTCAAGAAGATTTCTAATGCAAGTTTGTAGTTGTTATACCAATATGCTGAGCTTTTA	600
Qy	1995	gtgtagg 2001	
Db	601	ATGGTGG 607	
RESULT 15			
BI262492/c			
LOCUS 602953767t1 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5087992 3',			
DEFINITION mRNA sequence.			
ACCESSION BI262492			
VERSION BI262492.1 GI:14822752			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1 (bases 1 to 582)			
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.			
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL Unpublished (1999)			
COMMENT Contact: Robert Strausberg, Ph.D.			
Email: cgapb-femail.nih.gov			
Tissue Procurement: Dr. Louis Staudt, M.D., Ph.D. cDNA Library			
Preparation: Ling Hong/Rubin Laboratory			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Incyte Genomics, Inc.			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			

Plate: LLCMI846 row: e column: 17		High quality sequence start: 27		High quality sequence stop: 580.	
FEATURES		Location/Qualifiers		source	
1. 582		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="IMAGE:5087992"			
		/clone_lib="NIH_MGC_99"			
		/tissue_type="lymphoma, cell line"			
		/lab_host="DH10B (phage-resistant)"			
		/note="Organ: lymph; Vector: pOTB7; site_1: XhoI; site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."			
BASE COUNT		142 a	161 c	133 g	146 t
ORIGIN					
Query Match		27.4%;	Score 555;	DB 10;	Length 582;
Best Local Similarity		99.0%;	Pred. No. 2.5e-100;		
Matches 569;		Conservative 0;	Mismatches 5;	Indels 1;	Gaps 1;
Qy	779	tgagcagcagggcccgctctacgtctctcagcgtcaatggagtggtggtttgtg	838		
Db	582	TGAGCAGCAAGGGCCCGCTCTACCTGCTCTTTCAGCGTCAATGGGAGTGGGATTTTGTG	523		
Qy	839	gggtggcagataaagtcctccctgactgacagcaccagtcgggtctggtctcagg	898		
Db	522	GGGTGGCGGAGATGAAGTCCCGTGGACTACGGCACAGTCCCGGGTCTGGTCTCAGG	463		
Qy	899	acaagtgaagggaagtttgatgtccagtggtttttttaaggatgtacccaataacc	958		
Db	462	ACAAGTGAAGGGGAAGTTTGATGTCCAGTGGATTTTGTAAAGGATGTACCCAATAACC	403		
Qy	959	agctcccgccacatcaggtggagataaacgacacaaaccggtcacaaactcccgggaca	1018		
Db	402	AGCTCCCGGCACATCAGGCTGGAGAAATACGACAAACAAACCCTCCGCGGACA	343		
Qy	1019	cccaggagggtcccttagaaaaagcccaagtcgtgctgaaattatcagttctccacaagc	1078		
Db	342	CCCAGGAGGTGCCCCCTTAGAAAAAGCCCAAGCAAGTCTGTAATAATTATCAGTCTCTCAACAGC	283		
Qy	1079	acacaaacctccatcttcgacgactttgtctactacgagaagcgcca-gaggaggaggagg	1137		
Db	282	ACACAACCTCCATCTTCGACGACTTTGCTCACTACGAGAAGCGCCAGGAGGAGGAGG	223		
Qy	1138	tgggtgcaggaagcagcagagtcgaaacaaacaataatgaggcgcaaccggttcttaccatg	1197		
Db	222	TGGTGCAGCAAGGAACGGCAGAGTCTGAACAAACAATGAGGGCAACCACTTCTTACATG	163		
Qy	1198	tcttaacgttttgacttgaaacacagtttaaacacagtcgtctgtcagctccagtgctg	1257		
Db	162	TTCTTAACGTTTACATCTTTGAAAAACAGTTTAAAAACAGTGTGCTTGTGTAGCTCCAGCTGTGT	103		
Qy	1258	cgctcccggtgggggttgagtggttgcattcttgccttcttctgtgtgatttttgccag	1317		
Db	102	CGTCCCGTGGCGGGTGTGAGTGTTCATCTTTGCTTTTCTTGTGCTGATTTTGTGCCAG	43		
Qy	1318	atggaatcgaatttatttggctactttttctctat	1352		
Db	42	ATGATCTGCATTTTATTGTTGTAATTTCTCGGATGTAT	8		

Search completed: July 29, 2002, 17:18:42
Job time: 8478 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2028	100.0	2028	2	US-08-933-750C-75	Sequence 75, Appl
2	2028	100.0	2028	3	US-09-234-613-75	Sequence 75, Appl
3	135.4	6.7	2675	1	US-08-232-079-1	Sequence 1, Appl
C 4	53.8	2.7	19124	2	US-08-487-826B-13	Sequence 13, Appl
5	51	2.5	7218	1	US-08-233-463-14	Sequence 14, Appl
C 6	50.6	2.5	6243	2	US-09-056-075-1	Sequence 1, Appl
7	45.6	2.2	615	4	US-08-998-416-186	Sequence 186, App
8	45.2	2.2	858	4	US-08-998-416-595	Sequence 595, App
9	45.2	2.2	1620	2	US-08-461-775-10	Sequence 10, Appl
10	45.2	2.2	1620	3	US-09-031-606-10	Sequence 10, Appl
11	45.2	2.2	2668	2	US-08-461-775-11	Sequence 11, Appl
12	45.2	2.2	2668	3	US-09-031-606-11	Sequence 11, Appl
C 13	45	2.2	660	1	US-07-991-867B-32	Sequence 32, Appl
C 14	45	2.2	660	1	US-08-107-755A-32	Sequence 32, Appl
C 15	45	2.2	660	2	US-08-544-332-32	Sequence 32, Appl
C 16	45	2.2	1511	1	US-07-991-867B-8	Sequence 8, Appl
C 17	45	2.2	1511	1	US-08-107-755A-8	Sequence 8, Appl
C 18	45	2.2	1511	2	US-08-544-332-8	Sequence 8, Appl
C 19	45	2.2	4810	3	US-08-852-629-11	Sequence 11, Appl
C 20	45	2.2	4838	3	US-08-852-629-15	Sequence 15, Appl
C 21	44.8	2.2	6124	4	US-08-213-419B-3	Sequence 3, Appl
C 22	43.4	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
C 23	42.4	2.1	636	4	US-08-998-416-1137	Sequence 1137, App
24	42.4	2.1	837	4	US-08-998-416-288	Sequence 288, App
C 25	42.4	2.1	51952	3	US-08-947-823-1	Sequence 1, Appl
26	42.2	2.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
27	42	2.1	2576	1	US-08-471-033-35	Sequence 35, Appl

CLONE: 1573677
US-08-933-750C-75

Query Match	100.0%	Score	2028;	DB	2;	Length	2028;		
Best Local Similarity	100.0%	Pred.	No. 0;						
Matches	2028;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	caaaaggacagataaataagatacaaaatggttcgttatcatcagaagatcacagtccaag	60						
Db	1	CAAAAGGACAGATANTAAAGTACAAAATGGTTCGTTACATCAGAAGGATACAGTTCAATG	60						
Qy	61	acaatgactttgagccctacaccttactgagacagtcacaaatcagagtaaacagttaccctcaa	120						
Db	61	ACAATGACTTTGAGCCCTACCTTACTTGGCAGCTCAAAATCAGAGTAACAGTTACCCCTCAA	120						
Qy	121	tgagcgaaacctacactgtccagactattaccgcgcgtccatggatttcccttactccctca	180						
Db	121	TGAGCGACCCCTACCTGTGCCAGCTATTACCCGCGCTCCATTGGATTTCCTTACTTCCCTCA	180						
Qy	181	atgaggtcccggtgtctactgcaggggacctccgattccatcactcaaccacctaacggac	240						
Db	181	ATGAGGCTCCGTGFTACTTCAGGGGACCCTCCGATTTCCATACCTCACACCTACGGAC	240						
Qy	241	agctcagtaacggagaccatcattttaagcaagatgctgttttgggcagcctcgggggcc	300						
Db	241	AGCTCAGTAACGGGAGACCATCATTTTATGACAGATGCTGTTTTTGGGCAGCGCTGGGGCC	300						
Qy	301	tggggacaacatctatcagcacaggttcaatttttccctgaaaccctcgttctcag	360						
Db	301	TGGGGAACAACATCTATTCAGCACAGTTTCAATTTTTTCCCTGAAACCCTGCGTTCTCAG	360						
Qy	361	catgggggacaagtgggttctcaagtcagcagacccagagctcagcgtctcccagcacag	420						
Db	361	CATGGGGACAAAGTGGGTCTCAAGTCAAGCAGCAGACCCAGAGCTCAGCCTCTCCAGCACAG	420						
Qy	421	ccccagcttttgctcaaccgcagtatcagagccctcagcgaccccccagaccgcgtgg	480						
Db	421	CCCCAGCTTTTGGCTCAACCCGAGTATCAGAGCCCTCAGCAGCCACCCGACAGCCGCTGG	480						
Qy	481	gttgccccacgcacagaaaacgcggtcttgggcagagcgaggggctggcgagcgatagc	540						
Db	481	GTTGCCCCACGCACAGAAAACGCGGCTTTGGGCAGACGCGAGGGGCTGGCGCGATAGC	540						
Qy	541	aactctcctgaaacgctccagcctaattctctgccccagcgtcgaaatcccccgcgtcctt	600						
Db	541	AACCTCTCTGGAACGTCACGCTCAATTTCTGCCCCAGCTCGAATCCACCCGCTCCTT	600						
Qy	601	gaaatactgaagctgtcacagctacacccgaaaagatttgagtggaatctcaaaagc	660						
Db	601	GAAAACHTGAAGGCTGTCTACAGCTACAAACCCGAAGAGTTTGAGTGGAAATCTGAAAAGC	660						
Qy	661	gggcgtgtgttcatacataagagctactctgagcagacatccaccgcctccattaaagtac	720						
Db	661	GGGCGTGTGTTTCATCATCAAGAGTACTCTGAGGAGGACATCCACCGCTCCATTAAGTAC	720						
Qy	721	tccatctgggttagcacagagcaaggcaaacggccttgagcagcgcttcgcgtgcatg	780						
Db	721	TCCATCTGCTGTAGCACAGACGGCAACAAAGCGCTTGGACAGCGCTTCGGTGCATG	780						
Qy	781	agcagcagggggccgctcactcgtctctcagcgtcaatggagtggggcatctttgtggg	840						
Db	781	AGCAGCAAGGGGCCGCTACTCTCTTCAGCGTCAATGGGAGTGGGCATTTTGTGGG	840						
Qy	841	gtgcccagataagaagtcctcccgtagctacagcagccagtgccgggggtctggtctcaggac	900						
Db	841	GTGCCCCAGATGAAGTCCCGCGTACGCTACGGCACACAGTGCCTGGCGGTCTGGTCTCAGGAC	900						
Qy	901	aagtggaaaggggaagtttgatgtccagtggaatttttgaaggatgtaccacaataaacag	960						
Db	901	AAGTGGAAAGGGGAAGTTTGATGTCCAGTGGATTTTTTGTTAAGGATGTATCCCAATAACCCAG	960						
Qy	961	ctcgggcacatacaggtctgggagaataaacagcaacaaaccggtcacaacactccgggacacc	1020						

QY	1288	ttgcctctctctgttgatttttgcgcagatcgatctgcatttatctgtacattttctca	1347
Db	15968	TTAAATTTTATTTTATTTTTTTTATTAATTAATTTTTTATTTTTTTTATTCATTAA	15909
QY	1348	tgtattataactctgtagaagcactcaataaaaggagatttttttgcgcagcttatcaa	1407
Db	15908	ATTTTTTTAAATTTTTTTTTTATTAATTAATTTTTTATTAATTTTTTTTATTTAA	15849
QY	1408	tcagactgatctaagtgaagaatgcctataaaacaaagcatctattttgcagaga	1467
Db	15848	TTAAATTTTTTTTATTTATTTTTTTTTTTTATTTAAATAAATTTTTTTTATTTATGTATA	15789
QY	1468	aattgtgtcttaa----attcagtcatttgatattctgtgagacttcatactctcat	1522
Db	15788	TATTTTTTTTTTAACATTTTTTTTTTAAATTTTTTTTTTATTTATGATATATATTTTATTTTA	15729
QY	1523	ccctttattgctcttttagcaacataagaaccatgactgatttttgcatttttagagattt	1582
Db	15728	ATATATTTTTTCTTTTTTTTTTTCTTTTTTATGATATATATTTTTTTTTTTTAAAGTTTTT	15669
QY	1583	ctgataaaatcctctgaaaaactcgaaatacaaaaggcttaatgatatttttgcctatctga	1642
Db	15668	TTTTTTTCTCTCTCTGTTTTTATTTTTTTTTTATAAA--TCATTTTTTTTTTATATAAAATTT	15611

FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: pT3gpt-F15
 19-08-232-463-14

[illegible]

Query Match 2.2%; Score 45.2; DB 3; Length 2668;
Best Local Similarity 54.1%; Pred. No. 0.021;
Matches 92; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy	802	ctgctcttcagcgtcaatgggaagtgggaatttttctgggggtggccagatgaagaatcccc	861
Db	2169	GCGGTGCTGGCCGCCGCCGCCGTCGAGCGCGTGCCTCGATCGCGAGACGCCGGCCTC	2228
Qy	862	gtgaactacgcgccaccagtgccggggtctggtctcaggacaagtggaaagg	911
Db	2229	GAGGGCTACTCATATCAACCACCAAGTGCGCGAGCTCGACAAAGGGCCACGG	2278

REFERENCE: Gaurid, Michael E.
 TITLE OF INVENTION: No. 5476781e1 Entomopoxvirus Expression System
 NUMBER OF SEQUENCES: 66
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David R. Saliwanchik
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/991,867B
 FILING DATE: 12-DEC-1992
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO 92/14818
 FILING DATE: 12-FEB-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/827,685
 FILING DATE: 30-JAN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/657,584
 FILING DATE: 19-FEB-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Saliwanchik, David R.
 REGISTRATION NUMBER: 31,794
 REFERENCE/DOCKET NUMBER: UF114.C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 904-375-8100
 TELEFAX: 904-372-5800
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 660 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)


```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: PatentIn Release #1.0, Version #1.25
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/544.332
5 FILING DATE:
6 CLASSIFICATION: 435
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: US 07/991,867
9 FILING DATE: 07-DEC-1992
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 08/107,755
12 FILING DATE: 19-AUG-1993
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: WO 92/14818
15 FILING DATE: 12-FEB-1992
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 07/827,685
18 FILING DATE: 30-JAN-1992
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 07/657,584
21 FILING DATE: 19-FEB-1991
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Bencen, Gerard H.
24 REGISTRATION NUMBER: 35,746
25 REFERENCE/DOCKET NUMBER: UFI14.C4
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 904-375-8100
28 TELEFAX: 904-372-5800
29 INFORMATION FOR SEQ ID NO: 32:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 660 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: unknown
35 MOLECULE TYPE: DNA (genomic)
36 US-08-544-332-32

```

Query Match	2.2%	Score 45;	DB 2;	Length 660;
Best Local Similarity	46.3%;	Pred. No. 0.011;		
Matches 220;	Conservative 0;	Mismatches 250;	Indels 5;	Gaps 2;
QY 1293	ttctctgctgattgatttttccacagatggatctgcacatttatttgcacatttctctatgat	1352		
DB				
QY 512	TTTAAATTAATTAATTTTCTATAAATTTATAATCTTTTATTATTGATTGCATATTT	453		
DB				
QY 1353	tataatctctagaagtcactaataaaggagatttttttctcagcttcatcaatcaga	1412		
DB				
QY 452	AACCTCAACTATACATAAGGAAACACATTATTAAGTTACCAATTTATTTTTAGATATT	393		
DB				
QY 1413	ctgctctaattggaattgaagtatcccttaaaacaagcatctattttggcagaaattg	1472		
DB				
QY 392	ATTAATTTTTTTTAAATTTACTAAATTTATAAATAATATAGTCATTTATATTACATGAT	333		
DB				
QY 1473	tgctcttaaatcagtcatttgatattctctgagacattcatattctcattccctttattg	1532		
DB				
QY 332	TCACAATTTAAAAATCTATAGAAATGGTAGTATAATATTACTTATATTGCTATTTTGTG	273		
DB				
QY 1533	cttttttagcaaacataagaaccatgagtcattttgctcatttagagagtattctgataaaat	1592		
DB				
QY 272	TTATAGATATATCTAAATATGTTATATTTTAAATTTTGTATAAAAATTTAAATTAATA	213		
DB				
QY 1593	ctcttgaaataactgaaatcaaaaggctaatgatttttctgctattctgatttgcattt	1652		
DB				
QY 212	ATATTTTAAATTT--TGAATATATATAACTTTTAAATATTTTCTGGAAATATATTATTTAAATAAT	155		
DB				
QY 1653	tattattctgtattcggctctaaagtctcaatttaccacatttgattttctctgtagacagat	1712		
DB				
QY 154	TATTATCATATAATATATA--TGCAATTCCTCAATTAACATAATTTTTTAAATATAT	98		
DB				
QY 1713	aacttttaatttttcaaaattggcagacacatttttttttttttttttttttttttttttttt	1767		
DB				
QY 97	TAATATTAATTAACATTTATCTGTTTTATATATTTTTTTTTTAAATTTAATATATTTT	43		
DB				

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 15:06:14 : Search time 2520.46 Seconds
(without alignments)
16837.806 Million cell updates/sec

Title: US-09-877-633-2
Perfect score: 2028
Sequence: 1 caaagagcaagataataa.....agtacagacagtgtgatgga 2028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	2028	100.0	2028	6	AR116045	Sequence
2	1595	78.6	2615	9	AB055518	Homo sapi
3	1595	78.6	2952	9	BC003681	Homo sapi
4	1593.4	78.6	2429	9	AK000398	Homo sapi
5	1590.8	78.4	2695	9	BC016920	Homo sapi
c 6	870	42.9	140152	9	HSJ963E22	Human DNA
7	549.2	27.1	168012	2	AC025565	Homo sapi
8	549.2	27.1	172451	2	AC106748	Homo sapi
c 9	365.8	18.0	192994	2	AC011978	Homo sapi
10	341.6	16.8	2309	6	BC002559	Homo sapi
11	341.6	16.8	2750	6	AX048109	Sequence
12	340.6	16.8	2070	9	AF155095	Homo sapi
13	340.6	16.8	2730	9	AF192968	Homo sapi
14	340.2	16.8	2079	10	BC014797	Mus muscu
c 15	333.4	16.4	130984	2	AC022766	Homo sapi
16	333.4	16.4	162532	2	AL645729	Homo sapi
17	328	16.2	8533	9	AK024546	Homo sapi
c 18	321	15.8	163627	2	AL672026	Mus muscu
c 19	321	15.8	194985	10	AC002406	Mouse chr
c 20	318.2	15.7	141372	2	AC011779	Homo sapi
21	318.2	15.7	196319	9	CNS01DT4	Human chr
22	299	14.7	1573	10	BC022932	Mus muscu
23	271.8	13.4	2907	9	AK056219	Homo sapi
c 24	258.4	12.7	65091	2	AC100199	Mus muscu
25	226	11.1	379	6	AX072124	Sequence
c 26	214.6	10.6	32785	2	AC013560	Homo sapi
c 27	209.6	10.3	2679	3	AY069675	Drosophil
c 28	195.4	9.6	113474	2	AC013932	Drosophil
c 29	195.4	9.6	162921	3	AC007853	Drosophil
c 30	195.4	9.6	181132	3	AC008206	Drosophil
c 31	195.4	9.6	227398	3	AE003750	Drosophil
c 32	169	8.3	36384	9	AC004178	Homo sapi
33	169	8.3	104423	2	HSJ824K2	Human DNA
34	169	8.3	112578	9	AL645927	Human DNA
c 35	169	8.3	114868	9	HS994E9	Human DNA
c 36	169	8.3	129567	2	AL671618	Homo sapi
37	169	8.3	133836	9	AL662869	Human DNA
38	147.4	7.3	1840	8	AY072018	Arabidops
39	136.6	6.7	1768	8	AY056421	Arabidops
40	136.6	6.7	1836	8	AY065413	Arabidops
41	135.4	6.7	2675	6	118667	Sequence 1
c 42	132.8	6.5	423	6	AX330380	Sequence
43	132.4	6.5	2163	8	AF326910	Arabidops
44	125.8	6.2	2484	8	AY050425	Arabidops
c 45	119	5.9	101589	2	AP003897	Oryza sat

ALIGNMENTS

RESULT 1	AR116045	AR116045	Sequence 75 from patent US 6132973.	2028 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	AR116045	Sequence 75 from patent US 6132973.					
DEFINITION	AR116045	Sequence 75 from patent US 6132973.					
ACCESSION	AR116045	Sequence 75 from patent US 6132973.					
VERSION	AR116045.1	GI:14096367					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 2028)						
AUTHORS	Lal, P., Hillman, J.L., Bandman, O., Shah, P., Au-Young, J., Yue, H., Guegler, K.J. and Corley, N.C.						
TITLE	Human regulatory molecules						
JOURNAL	Patent: US 6132973-A 75 17-OCT-2000;						
FEATURES	Location/Qualifiers						
source	1. .2028						
BASE COUNT	520 a 478 c 463 g 567 t						
ORIGIN	/organism="unknown"						
Query Match	100.0%;	Score 2028;	DB 6;	Length 2028;			

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2028; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	caaaagcaagataataaagtacaaaatggttcattacatcagaagatacagttcatg	60
Db	1	CAAAAGGCAAGATAAATAAGTACAAAATGGTTCGTATACATCAGAAGGATACAGTTCAATG	60
Qy	61	acaatgactttgagccctacttactgacagtcacaaacagagtaacagttaccctcaaa	120
Db	61	ACAATGACTTTGAGCCCTACCTTACTGGACAGTCAAAATCAGAGTAACAGTTACCCCTCAA	120
Qy	121	tgagcgaacccctacctgtctcagctattaccgcctccattggatttccctaccctca	180
Db	121	TGAGCGACCCCTACCTGTCTCCAGCTATTACCCGCCCTCCATTGGATTTCCTTACTCCCTCA	180
Qy	181	atgaggtccgtggtctactcagagggaccctccgattcccatcaccctaccagcag	240
Db	181	ATGAGGCTCCGGTGTCTACTGCAGGGGACCCCTCCGATTCCATACCTCACCACTACGGAC	240
Qy	241	agtcagtaacggagaccatcatcttatgacagatgctgtttttgggcagcctggggcc	300
Db	241	AGCTCAGTAACGGAGACCATCATTTATGCACAGTCTCTTTTGGGCAGCCTGGGGCC	300
Qy	301	tggggaacaaactctatcagcacaggtttoaatttttccctgaaacccctggttctcag	360
Db	301	TGGGGAACAACATCTATCAGCACAGGTTCAATTTTTCCTGAAACCCCTGCGTCTCAG	360
Qy	361	catggggacaaagtgggtctcaagtacagacacccagcagctccaccctccacacag	420
Db	361	CATGGGGACAAGTGGGTCTCAAGGTCTACAGACCCAGAGCTCAGCCTCTCCACACACAG	420
Qy	421	ccccagctttggtcacaacggcaglatcagagccctcagcagcccccagacccgctgg	480
Db	421	COCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAGCAGCCACCACAGACCCGCTG	480
Qy	481	gtgccccacgcacacagaaacggcggtttgggcagcagcggggcgtggcgagatagc	540
Db	481	GTGCCCCACGCAACAGAAACCGCGGTTTGGGCAGAGCGGAGGGGCTGGCAGCGATAGC	540
Qy	541	aactctctggaaagctcagcctaatctgcgccagcgtcgaaatccaccacctcctt	600
Db	541	AACTCTCTGGAAAGCTCAGCGCTAAATCTGCCGCCAGCGTCAATCCACCCCGCTCTT	600
Qy	601	gaaactgaaggtgctcacagctacacacccgaagagtttgagtgaaatctgaaagc	660
Db	601	GAAAAACTGAAGGCTGCTCACAGCTACAAACCCGAAAGAGTTTGAGTGAATCTGAAAGC	660
Qy	661	gggcgtgtgtcatcatcaagagctactctgagggcagacatccacgcgtccaatgaagta	720
Db	661	GGGCGTGTGTTCATCATCAAGAGCTACTCTGAGGACGACATCCACCGCTCCCATTAAGTAC	720
Qy	721	tcacatcgtgtgtagcacagcagcgaacacagcgcttgacagcgccctccgctgcatg	780
Db	721	TCCATCTGCTGTAGCACAGACACGGCAACAGCGCGCTGGACAGCGCCTTCCCGCTGCATG	780
Qy	781	agcagaagggggccgctcactgctcttcagcgtcaaatgggagtgggcatttttgagg	840
Db	781	AGCAGAAGGGGCCGCTGTACTGCTCTTCAGCGTCAATGGGAGTGGGCATTTTGTGGG	840
Qy	841	gtggccgagatgaagtcccccgtagactacggcacacagtgccggggtctggctcaggag	900
Db	841	GTGGCCGAGATGAAGTCCCGCGTAGCTACGGCACAGTGC CGGGGCTGTGGTCTCAGGAC	900
Qy	901	aagtggaaaggggaagtttgatctcagtggaatttttgaagtgatgcccaataaccag	960
Db	901	AAGTGAAGGGGAAGTTTGATGTCAGTGGATTTTGTGTTAAGGATGTACCCAATAACCCAG	960
Qy	961	ctccggcacatcaggctggagaaataacacaaacacccggttcacaaactccccggacac	1020
Db	961	CTCCGGCACATCAGGCTGGAGATAACGACAACAAACCGGTCACAAACTCCCGGGACACC	1020
Qy	1021	caggagggtcccttagaaaaacccaagcgaagtgtgaaataattacagttccctacagac	1080

Db	1021	CAGGAGTGCCCTTATGAAAAAGCCAAAGCAAGTGCTGAAATATATCATCTCTCAACAGCAC	1080
Qy	1081	acaactccatcttcagagactttgctcactacgagaaagcgccagagaggaggtgag	1140
Db	1081	ACAACTCCATCTTCAGCAGCTTTGCTCACTACGAGAAAGCCGACAGAGGAGGAGGTGG	1140
Qy	1141	tgccaaagaaagcggcagagtcgaaacaaacaaatgagggcgaaaccagtttcttaccatgttc	1200
Db	1141	TGCCAAAGGAACGGCAGAGTCGAAACAAACAAATGAGGGGGAACCACTTCTTACATGTTTC	1200
Qy	1201	taacgtttgactttgaaacacagtttaaaacacagtcgtgtggtcagctccagtgctgct	1260
Db	1201	TAACTGTTGACTTTGAAACACAGTTTAAACACAGTGTGCTTGGTTCAGCTCAGTGTGCTCT	1260
Qy	1261	cccgctcggggggtcagtgctcactcttgccttctctgctgcttcttcttccacagag	1320
Db	1261	CCCGTCCGGGGGTGAGTGTTCATCTTTGCCCTTCTTCTGCTGTTGATTTTGGCCAGATG	1320
Qy	1321	gactgcatattttgtactttttctatgtattataaactcctgtagaugtcactaaataag	1380
Db	1321	GATCTGCATTTATTTGTACTTTTCTATGTATTAATCTCTAGAAAGTCACATAATAAG	1380
Qy	1381	gagtattttttgcagcttatacaatcagactgatcctaatactgaaatgtaagtatcct	1440
Db	1381	GAGTATTTTGTGTCAGCTTATCAATCAGACTGATCTAAATGTGAATGTAAATGTAATCCT	1440
Qy	1441	taaaacaaagcatctattttggcagaaatgtgttctttaaattcagtcatttgatcttc	1500
Db	1441	TAAAAACAAGCATCTATTTTGGCAGAAATGTGTCTTAAATTTCAATCTATTTGATATTTC	1500
Qy	1501	tgtgagacttcataattctcatcccttatttgccttttttagcaaaacataagaacacatgag	1560
Db	1501	TGTGAGACTTCATATTTCATCCCTTTATTCCTTTTAGCAAAACATAAGAAACCATGAG	1560
Qy	1561	tcatttgcatttgagagtatctctgataaaaactctctgaaataactgaaatacaaaaggtc	1620
Db	1561	TCATTTTGTCTATTTAGATTTCTGATAAAATCTCTGAAATACTGAAATCAAAAGGTT	1620
Qy	1621	aatgatcttttgcattctgctgcttcttcttcttcttcttcttcttcttcttcttcttctt	1680
Db	1621	AATGATTTTGTCTCATTTCTGATTTGTCATTTTATATCTGTTATCGGCTTAAAGTGCTTA	1680
Qy	1681	atttaaccatttgatttttctctgtagacagataaactttttaaattttttaaatttttggcagac	1740
Db	1681	ATTTACCCATTTGATTTTCTGCTAGACAGATAAATCTTTAAATTTTCAAAATTTGGCAGAC	1740
Qy	1741	acttttttttttttttttggaaatcttcttccagatctgttgcacactgaaacagccac	1800
Db	1741	ACTTTTGTGAAATCTTTTCTTCCAGATCTGTTGCCACCTGAAACAGCCACC	1800
Qy	1801	cgccctcactgctcctgggtgctcggattgggctggatgggtgttggggcatgagtgtgag	1860
Db	1801	CGTCCCTCACTGCTCCTGGTGTCCGATTTGGCTGGATGGTGTGGGGCATGATGTGTGGAG	1860
Qy	1861	gaactggaggtgctcttaggtctggttcaggggtcgggcattcttcttcttcttcttcttcttct	1920
Db	1861	GAACTGGAAAGGTGCTTAGGTCTGGTTCAGGGTGGGGCATCTCTTGTCTTGTGGCATCTCT	1920
Qy	1921	tttttaatttttaacacctttttttaaagaattctaaatgcctctttaaagttttttatatacaata	1980
Db	1921	TTTTTAAATTTTACACCTTTTCTTAAAGAAATCTAAATGCCGCTCTTAAGTTTTTATACCAATA	1980
Qy	1981	atgtgagctttaaagtgtaggtctggttagtcagacagctgtgagtga 2028	
Db	1981	ATGCTGAGCTTTAAGTGTAGGATCTGGTAGTACAGACAGCTGTGATGGA 2028	

RESULT	2
AB055518	
LOCUS	2615 bp mRNA linear PRI 07-AUG-2001
DEFINITION	Homo sapiens DACA-1 mRNA for dermatomyositis is associated with cancer
ACCESSION	AB055518
	putative autoantigen-1, partial cds.


```
QY 1898 cattcttgggtggtgcatctcttttaataattttacacaccttttcttaagaatacttaatgc 1957
|||||
Db 2038 CATCTCTTGTGTCACATCTTTTAAATTTTACACCTTTCTTAAGAATCTAATGC 2097
|||||

QY 1958 cgtcttaagtatttataccaataatgctgagcttttaagtgtaggatctggtagtagacag 2017
|||||
Db 2098 CGTCTTAAAGTTTATATACCAATAATGCTGAGCTTTAAAGTGTAGGATCTGCTAGTACAGAC 2157
|||||

QY 2018 agtgtagataga 2028
|||||
Db 2158 AGTGTGATGGA 2168
|||||

RESULT 3
BC003681
LOCUS BC003681 Homo sapiens, clone IMAGE:3453235, mRNA, partial cds. PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3453235, mRNA, partial cds.
ACCESSION BC003681
VERSION BC003681.1 GI:13277545
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2952)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-ghgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 4 Row: 9 Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7020459.
FEATURES
source
1..2952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3453235"
/tissue_type="Placenta, chorilocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
<L..1648
/codon_start=2
/product="Unknown (protein for IMAGE:3453235)"
/protein_id="AAH03681.1"
/db_xref="GI:13277546"
/translation="KGQDNKVQNGSLHOKDTHVNDFFPYLTGOSNQSNYSMSDPY
LSSYPPIGGPYSINEAPWSTAGDPPIPLYTYQLSNGDHHFMDAVFGQPGUGLN
NIYQHRFFNPENAFSAWGTSGSQQQTQSAYSSYTPPSSLGGTVVDGPGUGLN
DTLSKAPGMSLEQMGVLKIGDVSSAVKTVGSVSVALTGVLGNGGTNNMPVS
KPTSWAATAKPAKPPKMTKSGPVMGGGLPPPIKHNMDIGTDWNGKGPVPKAPVQ
QAPSPOAAPOPOVAOPLPAPOAPALAOPOYQSPQPPOTRWVAPRNRNAAFQSGGAG
SDNSPGNVOPNSAFVSESHPVLEKLAHNSYNPKFEWNLKSGRVFLIKSYSEDDIH
RSIKYSIWCSTEGHNRKLDSPFCRMSKGPVYLLFSVNGSHGFCGVAEMKSPVDVGTG
AGVSDQDKWKGAFDQWIFVFDVPPNNQUNRHIRENNDNKPV7NRSRDTQDEVPLEKARQV
```

LK11SSYKHTT5IFDDFAHYEKREBEEVVRKERSRQK*

BASE COUNT
ORIGIN

758 a 715 c 700 g 779 t

Query Match 78.6%; Score 1595; DB 9; Length 2952;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1628; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

```
QY 400 gctcagcctctccacagcagccccccagcttttgctcaaacgcagtagtatcagagccctcag 459
|||||
Db 872 GCTCAGCCTCTCCACAGCACAGCCCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAG 931
|||||

QY 460 cagccacccacagccgcgtgggttgcceccacgcaacagaaacgcgggtttggcgagagc 519
|||||
Db 932 CAGCCACCCACAGCCCGCTGGGTGGCCCCACGCAACAGAAACCGCGCGTTGGGCGACAGC 991
|||||

QY 520 ggaggggctggcagcgatagcaactctctggaaaacgtccagcctaatctgtgccccagc 579
|||||
Db 992 GGAGGGGCTGGCAGCGATAGCAACTCTCCTTGAAACGTCCAGCCTAATTTCTGCCCCAGC 1051
|||||

QY 580 gtcgaatccacccccctcttgaaaaactgaaggctgctcacagctacaaccgaaagag 639
|||||
Db 1052 GTCGAATCCACCCCTCTTGAAAAACTGAAGCTGCTCACAGCTACAAACCCGAAGAG 1111
|||||

QY 640 tttagtggaatctgaaaagcggcggtgttctcatcaagagctactctgaggacgac 699
|||||
Db 1112 TTTGAGTGGAACTTGAAAGCGGCGGTGTTTCATCATCAAGAGCTACTCTGAGGAGGAC 1171
|||||

QY 700 atccacgctccatgaagtactccatctggtgtagcacagagcagcgcaacagcgctg 759
|||||
Db 1172 ATCCACCGCTCCATTAACTACTCCTGTTAGCACAGAGCAGCGCAACAGCGCGCTG 1231
|||||

QY 760 gacagcctctccgctgcatgagcagcagggggcccgctctactgctcttcacggtcaat 819
|||||
Db 1232 GACAGCGCTTCCGCTGCATGAGCAGCAGGGGCGCGCTCTACTGCTCTTCAGCGTCAAT 1291
|||||

QY 820 gggagtgggcatttttgggggtggcgagatgaagcccccgctggactacggcaccagt 879
|||||
Db 1292 GGGAGTGGGCATTTTGTGGGTGGCGAGATCAAGTCCCCCGCTGGACTACGCGACCACT 1351
|||||

QY 880 gccggggtctggtctcagagacaagtggaaagggaagtgtgatgtccagtggaattttgtt 939
|||||
Db 1352 GCCGGGTCTGCTCAGGACAAAGTGAAGGGGAAGTTTGATGTCCAGTGGATTTTGT 1411
|||||

QY 940 aagatgtaccacaaataacagctccgcacatcagctggagaaataacacacaacacccg 999
|||||
Db 1412 AAGATGTACCCCAATTAACAGCTCCGGCATCATCAGCTGGAGAAATACGCAACAACACCG 1471
|||||

QY 1000 gtcacaaactcccgagacacccagagggtgaccttagaaaaagccaagtcgtgaaa 1059
|||||
Db 1472 GTCACAAACTCCCGGACACCCAGAGGTGCCCTTAGAAAAAGCCAAGTGTGTGAAA 1531
|||||

QY 1060 attatcagttctacaaagacacacacacacacacacacacacacacacacacacacac 1119
|||||
Db 1532 ATTATCAGTTCCTACAAAGCACACAACCTCCATCTTCGACGACTTTGTCTACTACGAGAAG 1591
|||||

QY 1120 cgcga-gaggagagaggtgtgcgaaggaacgcagagtcgaacaacaacaataaggg 1178
|||||
Db 1592 CGCCAGGAGGAGGAGGAGGTGTGCGCAAGGAACGCGAGAGTGCNAACAACAATGAGGG 1651
|||||

QY 1179 cgaaccagttctctacatgttctaaagctttgactttgaaaacagtttaaacacgctgtgc 1238
|||||
Db 1652 CGAACCAAGTTTCTTACATGTTCTAACGTTTGACTTTGAAAACAGTTTAAACACGTTGTC 1711
|||||

QY 1239 ttggtcagctcccgctgtgtgtgtcccgctgccccggttgagtggtgcactcttgccttctt 1298
|||||
Db 1712 TTGGTCAGCTCCAGTGTGCTCCCGCTGCGGGGGTTGAGTGTGGTTCATCTTTGCTTTCTT 1771
|||||

QY 1299 gtcgttgatttttgcacagatgactgtcatttatgttacttttctatgtattataat 1358
|||||
Db 1772 GTCGTTGATTTTGGCCAGATGGATCTGCATTTATTTGTACTTTTCTTCTATGTATTATAAT 1831
|||||
```

```
Qy 1359 cctgtagaagtcactaataaaggaggtattttttttttgtcagcttactcaatcagactgac 1418
Db 1832 CCTGTAGAAGTCACATAATAAGGAGTA-TTTTTTTTGTGCAGCTTATCAATCAGACTGATC 1890
Qy 1419 taatgtgaatgttaagtatccttaaaacaaagcatctattttggcagaaaattgtttct 1478
Db 1891 TAAATGTGAATGTAAAGTATCCTTAAACAAAGCATCTATTTTGGCAGAAATGTGTCT 1950
Qy 1479 taaattcagtcatttgattctgtgagacttcatttctcattccctttatgtctttt 1538
Db 1951 TAAATTCAGTCATTGTATCTCTGTGAGACTTCATATTTCTCATCCCTTATTTGCTTTT 2010
Qy 1539 agcaacataagaacaccatgagtcattttgttcattagagattctctgataaaactcttg 1598
Db 2011 AGCAAACTAAGAAACCATGAGTCATTTTGTCAATTAGATTTCTCTAGACAGATAACTTT 2070
Qy 1599 aaaatactgaaatcaaaagggttaagtatttttttttctcattctgatttgctatttat 1658
Db 2071 AAAATACCTGAAATCAAAAGGTTAATGATTTTTTGTTCATCTGATTTGTCTATTTAT 2130
Qy 1659 ctgttaccgtgtcctaagtcctaatttaccatttgattttttctgtctagacagataacttt 1718
Db 2131 CTGTTATCGGCTCTAAGTGCTAAATTTACCCATTTGATTTTCTCTAGACAGATAACTTT 2190
Qy 1719 taattttcaaatgtggcagacac-tttttttttttttttttttttttttttttttttccagat 1777
Db 2191 TAAATTTTCAAAATTTGGCAGACACTTTTTTTTTTTTTTTTTTTTTTGTGAAATCTTCTTCCAGAT 2250
Qy 1778 ctgttgcccaatgaacagcaccgcctcactgcctcgtcgtcgtcgtcgtcgtcgtcgtcgt 1837
Db 2251 CTGTTGGCCACTGAACAGCCACCCTCCCTCACTGCTCCTGCTGCTGCTGCTGCTGCTGCTG 2310
Qy 1838 gtgttgaggcagatgtgtgaggaactggaagtcctttagtctgtcgtcgtcgtcgtcgtcgt 1897
Db 2311 GTGTTGGGCATGATGTGTGAGGAACCTGGAAGTGTCTTTAGTCTGCTTTCAGGTCGGG 2370
Qy 1898 catttttgttggcacatcttttttaaatctttacaccttttttaagaattcttaagtc 1957
Db 2371 CATCTTTTGTGTTGGCACATCTTTTAAATTTTACACCTTTTCTTAAAGATTTCTAATGC 2430
Qy 1958 cgtcttaagttttataccataatgctgagctttaagttaggatctggtagtagacagac 2017
Db 2431 CGTCTTAAGTTTATACCAATAATGCTGAGCTTTAAGTGTAGGATCTGGTAGTACAGAC 2490
Qy 2018 agtgtgatgga 2028
Db 2491 AGTGTGATGGA 2501

RESULT 4
AK000398 AK000398 2429 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20391 fis, clone KAI4640.
ACCESSION AK000398
VERSION AK000398.1 GI:7020459
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens ileal mucosa cDNA to mRNA, clone_lib:kaia
clone:KAI4640.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K.,
Hirao,M., Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
TITLE Unpublished (2000)
JOURNAL 2 (bases 1 to 2429)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
JOURNAL University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
```

```
COMMENT Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source Location/Qualifiers
1..2429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="KAI4640"
/tissue_type="ileal mucosa"
/clone_lib="kaia"
/note="cloning vector pWE18SFL3"
10..603
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:7020460"
/translation="MLEFLSLGANGTTSISGTSIFSLKLRSHQGGQVGLKVSRRPAP
RMAAATTPRAPVWVRLMGSQAFTATPPSRPPPIKHMMDIGTMDKNGPVKAPVPO
QAPSPQAAPQVQAQPLPAQPPALAQPOYQSPOQPOTRWVAPRNRNAAFGQSGGAG
SDSNSPGNVQPNAPSVEHPVLEKLAHSHYNKEF"
BASE COUNT 616 a 571 c 569 g 673 t
ORIGIN
Query Match 78.6%; Score 1593.4; DB 9; Length 2429;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1627; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
Qy 400 gctcagcctctccacgacacccccagcttttggctcaacgcagatatacagagccctcag 459
Db 358 GCTCAGCCTCTCCAGCAGCAGCCCGCAGCTTTGGTCAACCGCAGTATCAGAGCCCTCAG 417
Qy 460 cagccacccacgacccgctgggttggccacgacacaaacgcgcttgggagagc 519
Db 418 CAGCCACCCACAGACCCGCTGGTGGTCCCGCCACGCAACAAACCGCGCGCTTTGGGAGAGC 477
Qy 520 ggaggggctgagcagatagcaactctcctggaacgctccagcctaaattcttgcctccagc 579
Db 478 GGAGGGGCTGGCAGCGATAGCAACTCTCTCTGAAACGCTCCAGCCTAATTCGCCCCAGC 537
Qy 580 gtcgaatcccccctccttgaaaaactgaagctgctcagcagctacaacccgaaaagag 639
Db 538 GTCGAATCCACCCCGCTCTTTGAAAACTGAAGGCTGCTCAGAGCTACAACCCGAAAGAG 597
Qy 640 -tttgaagtgaatctgaaaagggcgctgtgttcctatcatcaagagctactctcgagcga 698
Db 598 TTTTGAGTGGAAATCTGAAAGCGGGCGTGTGTTTCATCATCAAGAGCTACTCTGAGAGCA 657
Qy 699 catccaccgctccatttaagtactcctcctgtgtgtagcagagcagcagcgaacaaagcct 758
Db 658 CATCCACCGCTCCATTAAGTACTCTCATCTGTGTAGCAGAGCAGCAGCAACAGCGCT 717
Qy 759 ggacagccttccgctcagcagcagcaagggggccctctacctgtctctcagcgtcaa 818
Db 718 GGACAGCCCTCTCCGCTGCATGAGCAGCAAGGGGCGCGCTTACCTGCTCTTCAGGCTCAA 777
Qy 819 tggagtggtgcatctttgtgggtggccgagatgaagtcctcccgctggactacggcaccag 878
Db 778 TGGGAGTGGGCAATTTTGTGGGTGGCCGAGATGAAGTGAAGTGTGATGTCAGTGCATTTTCT 837
Qy 879 tgccggggctctgggtctcaggacaagtggaaagggggaagtttggatgtccagtgattttgt 938
Db 838 TGCCGGGCTCTGGTCTCAGGACAAGTGAAGGGGAAGTTTGTATGTCCTCAGTGCATTTTCT 897
Qy 939 taagatgtacccaataaccagctccgcacatcaggtcgtggagataacgacaaacac 998
Db 898 TAAGGATGTACCAATAACCAAGCTCCGGCAGATCAGGCTGGGAATAACGACAAACACC 957
```

Qy 999 ggtcaaaactccggagacaccagaggtgaccttagaaaaagcgaagcaagtgctgaa 1058
|||||
Db 958 GGTCAAAACTCCGGGACACCAGGAGGTGCCCTTAGAAAAAGCCAAAGTGCTGAA 1017
Qy 1059 aattatcagttcctacaaagacacaaacctccatcttcgcagcagctttgctcaactaagaaa 1118
|||||
Db 1018 AATTATCAGTTCCTACAAGCACACAACCTCCATCTTCGACGACTTTGCTCACTACGAGAA 1077
Qy 1119 ggcaca-gagagagagaggtgtgctgcaagaaagcagagctgcaaaacaatgaagg 1177
|||||
Db 1078 GCGCCAGGAGGAGGAGGTGTGCGCAAGGAACGCGAGCTCGAAACAACAATGAGG 1137
Qy 1178 ggaacacagtttctacatgttcttaacggttgactttgaaacagtttaaaacacagtg 1237
|||||
Db 1138 GCGAACCACTTCTTACATGTTCTTAACGTTTGACTTTGAAACAGTTTAAACACAGTGTG 1197
Qy 1238 cttggtcagctccagtgctgcctccggtgcgggggtgagtgctgcatcttggctttct 1297
|||||
Db 1198 CTTGGTCACTCCAGTGTCTGCTCCGCTGCGGGGTGAGTGTTCATCTTTGCCCTTCT 1257
Qy 1298 tctgctgattttcccgagatgagctgcatttatttatactttttctatgtattataa 1357
|||||
Db 1258 TGTGCGTTGATTTTGGCCAGATGGATCTGCATTTATTTGTACTTTTCTATGTATTATTA 1317
Qy 1358 tctgttagaagtcactaaataaaggaggtatttttttgcagcttatacatcagactgat 1417
|||||
Db 1318 TCCTGTAGAAGTCACATAAAGGAGTA-TTTTTTTGTGCAGCTTATCAATCAGACTGAT 1376
Qy 1418 ctaatgtgaaagtgaagtaaccttaaaacaaagcatctattttggcagaaattggttcc 1477
|||||
Db 1377 CTAATGTGAATGTAAAGTATCTCTTAAACAAAGCATCTATTTTGGCAGAAATTGTGTTT 1436
Qy 1478 ttaaatcagtcatttgatctctgagagacttcatttccatccctttatttgccttt 1537
|||||
Db 1437 TTAATTCAGTCATTGTATTTCTGTGAGACTTCATTTTCATCTCCCTTATTTGCTTTT 1496
Qy 1538 tagcaacaataagaacacatagctatctttgtcatttagatattctgtataaatctctt 1597
|||||
Db 1497 TAGCAAAACATAAGAACCATGAGTCATTTTGTTCATTTAGAGATTTCTGTATAAATCTCTT 1556
Qy 1598 gaataactgaatacaataaagggttaatgatttttttgcattctgatttgcatttctat 1657
Db -1557 GAAATPACTGAATCAAAAGGTTAATGATTTTGTTCATTTCTGATTTCTCAATTTTATTA 1616
Qy 1658 tctgttactcgttcaagtgtcaatttacocattgtattttctctgctagacagataactt 1717
|||||
Db 1617 TCTGTTATCGGCTAAGTGCTAATTTACCCATTTGATTTTCTGCTAGACAGATAACTT 1676
Qy 1718 ttaatttttcaatttggcagacactttttttttttttttttttttttttttttttttccctccagat 1777
Db 1677 TTAATTTTCAATTTGGCAGACACTTTTTTTTTTTTTGTGAAATCTTTTCTTCCAGAT 1736
Qy 1778 cgtgtgccactgaacagacaccccgctccctcactgctcctggtgctcagatggtgctgagtg 1837
Db 1737 CTGTGCGGCACTGAACAGGCACTCGTCCGCTCACTGCTCGTGGTTCGCGATTGGGCTGGATG 1796
Qy 1838 gctgtgggcagatgctgtgaggaactggaaggtgacttttaggtctggttcagggctcggg 1897
Db 1797 GTGTGGGCGCATGATGTGTGGAGGAACCTGGAAGGTGCTTTAGTCTGTGGTTCAGGGTCCGGG 1856
Qy 1898 cattctttgtgttgacacatctttttaaattttacacctttttcttaagaatttctaatgc 1957
Db 1857 CATTCCTTTGTGTTGCACATCTTTTAAATTTTACACCTTTTCTTAAAGATTTCTAATGC 1916
Qy 1958 cgtcttaagtttttatacaacaataatgctgagctttaaagtgtaggatcctggttagtacagac 2017
Db 1917 CGCTTTAAAGTTTTTATACCAATAATGCTGAGCTTTAAAGTGTAGGATCTGGTAGTACAGAC 1976
Qy 2018 agtgatgagga 2028
Db 1977 AGTGTATGGA 1987

RESULT 5
BC016920
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
BC016920 Homo sapiens
2695 bp mRNA linear PRI 09-NOV-2001
Homo sapiens, clone IMAGE:3853372, mRNA, partial cds.
BC016920
BC016920.1 GI:16877330
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2695)
Straussberg, R.
Direct Submission
Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcdépaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 20 Row: 0 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8923363.
Location/Qualifiers
1..2695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3853372"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..1403
/codon_start=3
/product="Unknown (protein for IMAGE:3853372)"
/protein_id="AAH16920.1"
/db_xref="GI:16877331"
/translation="HASDHFMHDAVFGQGLGNNTYOHPRNFPFPPENPAFSAMGTSGS
GQOQVTKSVSSVYTPPSSLGCTVVDQPGPHSDTLKAPGMNSLEQGMVLKIGDV
SSSAQVTSVSVVLTGLVSGNGGTNNMPVSKPTSMAAITASKPAKPQPMKTKSG
PVMGGLPPPIKHNNDGTPNDKGPVPKAPYQQAAPQQAAPQQAAPQQAAPPA
LAQOYQSPQPPQTRWAPNRNNAFGSGGSDSPGNSPNSAPSVSHPVLE
KLKAHSYNPKFEWNLKSGRVFIISKYSEDDIHKRSIKYSIWCSTEHGKRLDSAPRC
NSSKCPVYLLFSVNGSHGCGVAEMKSPVDYCTAGVMSQDKKCKFDVQWLFVXKDV
NNQLHRLENNNDKPVNNSRDTQEVPLEKAKQVLKIISYKHTTIFDDPAHYEKRO
EEEEVRRKERQSRNQ"
BASE COUNT 674 a 644 c 654 g 723 t
ORIGIN
Query Match 78.4%; Score 1590.8; DB 9; Length 2695;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1625; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
Qy 400 gctcagcctctccacagccccagcttttgctcaccagcagtatcagagccctcag 459
|||||
Db 627 GCTCAGCCTCTCCACGACAGACGCCCCAGCTTTGGCTCAACCGCAGTATCAGAGCCCTCAG 686
Qy 460 cagccacccacagaccgctgggttgccccacgcaacagaaacgcggcgcttggcgagagc 519

Db 687 CAGCACCCAGACCGCTGGTTGCCCCACGCAACAAACGGCGCTTTGGCAGAGC 746
Qy 520 ggaggggctgagcagatagcaactctctctggaacgctccagcctaattctgccccagc 579
Db 747 GGAGGGCTGCAGGATAGCAACTCTCTGCTGAAACGCTCCAGCCTAATCTGCCCCAGC 806
Qy 580 gtcgaatcccccctctcttgaaaactgaaggctgctcacagctacaaccccaagag 639
Db 807 CTCGAATCCACCCCGCTCTTGA AAAACTGAAGCTGCTCACAGCTACAACCCGAAAGAG 866
Qy 640 tttaggtgaatcgaagaaggcgctgtgttcacatcaagaagactactcgaagacagc 699
Db 867 TTGAGTGAATCTGAAAGCGGCGTGTGTTCAATCAAGAGCTATTCTGAGGATGAC 926
Qy 700 atccacgcctccatgaagtaactcctctgtagcagagcagcgcaacaagcgctg 759
Db 927 ATCCACCGCTCCATTAAGTACTCATCTGTGTAGCAGACAGCGCAACAGCGCCTG 986
Qy 760 gacagcgcttccgctgcatgagcagcaaggccgctctacctgctcttcagcgtcaat 819
Db 987 GACAGCGCTTCCGCTGCATGAGCAGCAAGGGCGCGCTACCTGCTCTTCCAGCGTCAAT 1046
Qy 820 gggagtgggcaatttctggtgggtggcgagatgaagtcccccgctggactcagcagc 879
Db 1047 GGGAGTGGGCAATTTTGTGGGGTGGCGGAGATGAAGTCCCGCTGGACTACGGCACCACT 1106
Qy 880 gccgggctgagtcaggaacaagtgaagggaagtttgatgtccagtgattttgtt 939
Db 1107 GCCGGGGTCTGGTCTCAGGACAGTGGGAAGGGGAAGTTGATGTCCAGTGGATTTTGT 1166
Qy 940 aagatgtaccacaataaccagctccgacacatcaggtggagataacagcaacaacccg 999
Db 1167 AAGGATGTACCAATAACCACTCCGGCACATCAGGCTGGAGATAAGCAACAACACCG 1226
Qy 1000 gtcacaaactcccgggacaccagagagtgcccttagaaaaagccaagcagtgctgaaa 1059
Db 1227 GTCACAACTCCCGGGACACCGAGAGGTGCCCTTAGAAAAAGCAACAAAGTGTGAAA 1286
Qy 1060 attacagttctcaagcacacacacacctccatcttcagcagacttgcctcactagagag 1119
Db 1287 ATTATCAGTTCCTACAGCACACAACTCCATCTTCGACGACTTTGCTCCTACGAGAG 1346
Qy 1120 cgcca-gaggaggaggtggtgcgcaaggagcagcagagtcgaacaaacaatgaggg 1178
Db 1347 GCCCAGGAGGAGGAGGTGCTGCAGCAAGCAAGCGCAGAGTCGAACAAACAATGAGGG 1406
Qy 1179 cgaaccagttcttacatgttctaacgtttgactttgaaacagtttaaaacacagtgctg 1238
Db 1407 CGAACACAGTTTCTTACATGTCTTAACGTTTGAACAGTGTGAAAAACAGTGTGC 1466
Qy 1239 ttggtcagctcagtggtgctccgctgcgggttgagtggtgcatcttgcctcttctt 1298
Db 1467 TTGTCACCTCCAGTGTCTGCTCCGCTCGGGGTTGAGTGTGGATCTTTGCCCTTCTT 1526
Qy 1299 gtcgttgattttgcccagatgagatcatttttgaacttttttgaacttttctattataat 1358
Db 1527 GTCGTTGATTTTGGCCAGATGATCTGCATTTATTTGTACTTTTCTATGATTTATAAT 1586
Qy 1359 cctgtagaagtcactaataaggagatatttttttgcagcttatcaatcagactgac 1418
Db 1587 CCTGTAGAAGTCACTAATAAAGGAGTA-TTTTGTGTGTCAGCTTATCAATCAGACTGATC 1645
Qy 1419 taatgtaaatgaagtatccttaaaaaaagcatctatttttgccagaaaattgttct 1478
Db 1646 TAATGTGAATGTAAATATCTTAAACAAAGCATCTATTTTGGCAGAAATGTGTCT 1705
Qy 1479 taattcagctatttgatattctgtgagacttcatttctcatccctttattgttctttt 1538
Db 1706 TAAATTCAGTCATTGTATCTCTGTGAGACTTCATATTCTCATCCCTTTATTGCTTTT 1765
Qy 1539 agcaaacataagaaacacatgagtcatttttgcatttagatttctgataaatactcttg 1598
|||||

Db 1766 AGCAAAACATAAGAAACCATGAGTCAATTTGTGCTATTAGAGTATTCTGATAAAATCTCTTG 1825
Qy 1599 aaataactgaatacaaaaggttaagtattttttgttcattctctgatttgcattttattat 1658
Db 1826 AAATATCTCAAAATCAAAAGGTTAATGATTTTTTGTTCATCTGATTTGCTATTTAT 1885
Qy 1659 ctgttatcggtctaaagtgcataatttaccatttgatttttctctgctagacagataacttt 1718
Db 1886 CTGTTATCGGTCTAAAGTCTAAATTTACCCATTTGATTTTCTGCTAGACAGATAACTTT 1945
Qy 1719 taatttttcaaaatttggcagacacttttttttttttttttttttttttttttttttttttttt 1778
Db 1946 TAAATTTTCAAAATTTGGCAGACAC-TTTTFTTTTTTTTGA AAAATCTTTCTTCAGATC 2004
Qy 1779 tgttgcccaactgaacagcaccccgctccctcactgctccctgggtgcgattgggctggatgg 1838
Db 2005 TGTGGCCCACTGAACAGACCCCGTCCCTCACTGCTCTGCTGGTGTCCGATTTGGCTGGATGG 2064
Qy 1839 tgttggggcatgatgtgtggagggaactggaagggtctttagtctgtggttcagggtcgggc 1898
Db 2065 TGTGGGGCATGATGTGTGGAGAACTGGAAGGTGCTTTAGGTCTGTTCAGGGTCGGGC 2124
Qy 1899 attcttgttgttgcacatctttttaaattttacacottttcttaagaattctaatgcc 1958
Db 2125 ATCTTTTGTGTTGCACATCTTTTAAATTTTACACCTTTTCTTAAGAAATTTCTAATGCC 2184
Qy 1959 gtcctaaatttttataccaataatgctgagctttaagttaggtcagctcgttagtagacagaca 2018
Db 2185 GCTTTAAGTTTTTATACCAATAATGCTGAGCTTTAAGTGTAGGATCTGGTAGACAGACA 2244
Qy 2019 gttgtagtga 2028
Db 2245 GTGTGATGGA 2254
RESULT 6
HSJ963E22/c
LOCUS
DEFINITION Human DNA sequence from clone RP5-963E22 on chromosome 20 Contains the 3' end of a novel gene similar to NY-REN-2 Antigen, 5 Cpg islands, ESTs, STSS and GSSs, complete sequence.
ACCESSION AL096828.29 GI:9663381
VERSION AL096828.29
KEYWORDS HTG: Cpg island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140152)
AUTHORS Lloyd,D.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Aug 2, 2000 this sequence version replaced gi:958454.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RP5-963E22 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP5-963E22 is at 1 in this sequence. The true left end of clone RP11-261N11 is at 140053 in this sequence. The true right end of clone RP11-305P22 is at 53294 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-963E22 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES

Source

Location/Qualifiers

1..140152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-963E22"
/clone_lib="RPCI-5"

repeat_region
1..237
/note="AluY repeat: matches 3. .236 of consensus"
repeat_region
240..282
/note="Alu repeat: matches 1. .43 of consensus"
misc_feature
complement(750..1254)
/note="match: GSS: Em:AQ604429"
misc_feature
1432..1979
/note="match: STS: Em:HS58315T"
repeat_region
3187..3483
/note="AluX repeat: matches 1. .297 of consensus"
repeat_region
4567..4656
/note="MER53 repeat: matches 15. .104 of consensus"
repeat_region
4810..5072
/note="MLTID repeat: matches 116. .454 of consensus"
repeat_region
5073..5173
/note="L1MA3 repeat: matches 6200. .6300 of consensus"
misc_feature
complement(5093..5757)
/note="match: GSS: Em:AQ939378"
repeat_region
5174..5217
/note="MLTID repeat: matches 453. .496 of consensus"
misc_feature
complement(5569..5757)
/note="match: GSS: Em:AQ939335"
misc_feature
complement(5580..5718)
/note="match: GSS: Em:AQ939345"
misc_feature
complement(6984..7421)
/note="match: GSS: Em:AQ891633"
repeat_region
7582..7880
/note="AluSg repeat: matches 5. .302 of consensus"
repeat_region
8135..8178
/note="4 copies 11 mer 86% conserved"
repeat_region
8273..8352
/note="5 copies 16 mer 77% conserved"
repeat_region
8274..8345
/note="6 copies 12 mer 80% conserved"
repeat_region
8363..8462
/note="25 copies 4 mer gatg 77% conserved"
repeat_region
8366..8461
/note="6 copies 16 mer 80% conserved"
repeat_region
8367..8462
/note="8 copies 12 mer 78% conserved"
repeat_region
8369..8464
/note="3 copies 32 mer 86% conserved"
repeat_region
8484..8587
/note="26 copies 4 mer gatg 76% conserved"
repeat_region
8487..8582
/note="6 copies 16 mer 80% conserved"
repeat_region
8488..8583
/note="8 copies 12 mer 78% conserved"
repeat_region
8490..8585
/note="3 copies 32 mer 86% conserved"
repeat_region
8605..8672
/note="17 copies 4 mer gatg 79% conserved"

repeat_region
8608..8671
/note="4 copies 16 mer 84% conserved"
repeat_region
8609..8680
/note="6 copies 12 mer 77% conserved"
misc_feature
complement(10049..10339)
/note="match: GSS: Em:AQ201440"
misc_feature
10351..10851
/note="match: GSS: Em:AQ780882"
misc_feature
10502..10839
/note="match: GSS: Em:AQ147226"
repeat_region
11097..11236
/note="2 copies 70 mer 87% conserved"
repeat_region
11103..11230
/note="4 copies 32 mer 88% conserved"
repeat_region
11105..11242
/note="3 copies 46 mer 84% conserved"
repeat_region
11106..11227
/note="61 copies 2 mer ct 89% conserved"
repeat_region
11108..11239
/note="3 copies 44 mer 89% conserved"
repeat_region
11111..11230
/note="10 copies 12 mer 90% conserved"
repeat_region
11118..11237
/note="4 copies 30 mer 88% conserved"
repeat_region
11170..11303
/note="2 copies 67 mer 86% conserved"
repeat_region
11186..11275
/note="2 copies 45 mer 94% conserved"
repeat_region
11212..11337
/note="6 copies 21 mer 66% conserved"
repeat_region
11214..11334
/note="11 copies 11 mer 66% conserved"
repeat_region
11220..11336
/note="13 copies 9 mer cccctctcc 66% conserved"
misc_feature
complement(11220..11249)
/note="tandem repeat. Single clone region. Assembly confirmed by restriction enzyme digest data"
repeat_region
11224..11331
/note="3 copies 36 mer 81% conserved"
repeat_region
11228..11325
/note="49 copies 2 mer tc 81% conserved"
repeat_region
11233..11328
/note="3 copies 32 mer 86% conserved"
repeat_region
11239..11352
/note="3 copies 38 mer 78% conserved"
repeat_region
11243..11332
/note="3 copies 30 mer 84% conserved"
repeat_region
11255..11326
/note="4 copies 18 mer 84% conserved"
misc_feature
11390..11273
/note="CpG island"
/evidence=not_experimental
repeat_region
11397..11472
/note="19 copies 4 mer tgtg 69% conserved"
repeat_region
12829..13053
/note="5 copies 45 mer 74% conserved"
repeat_region
12834..13057
/note="7 copies 32 mer 65% conserved"
repeat_region
12835..13018
/note="4 copies 46 mer 69% conserved"
repeat_region
12836..13045
/note="3 copies 70 mer 76% conserved"
repeat_region
12840..13055
/note="12 copies 18 mer 66% conserved"
repeat_region
12847..13062
/note="6 copies 36 mer 67% conserved"
repeat_region
12849..13052
/note="12 copies 17 mer 65% conserved"
repeat_region
12851..13040
/note="5 copies 38 mer 73% conserved"
repeat_region
12891..13040
/note="5 copies 30 mer 68% conserved"
repeat_region
12936..13047

misc_feature /note="7 copies 16 mer 66% conserved" 13006

repeat_region /note="Tandem repeat. Forced join. Gap size estimated to be approximately 500bp by BAMHI and ECORI digest data." 13276..13338

repeat_region /note="3 copies 21 mer 85% conserved" 13844..13977

repeat_region /note="MR58 repeat: matches 92. .233 of consensus" 16177..16269

repeat_region /note="MLTLD repeat: matches 51. .146 of consensus" 16367..16629

repeat_region /note="MLTLD repeat: matches 206. .494 of consensus" 17818..18094

repeat_region /note="MLT1A1 repeat: matches 2. .278 of consensus" 18125..19582

repeat_region /note="L1M4 repeat: matches 3996. .5549 of consensus" 19583..19986

repeat_region /note="MSTA repeat: matches 1. .426 of consensus" 42.9%; Score 870; DB 9; Length 140152; Best Local Similarity 99.9%; Pred. No. 1.9e-190; Matches 881; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1147 agaacggcagtcgcaaaacaaatgagggaaccagtttcttatgttttaacgt 1206

Db 106290 AGGAACGGCAGCTCGAACAACAAATGAGGGCGAACAGTTCTTACATGTTCTTAACGT 106231

QY 1207 ttgaactttgaaacagtttaaaacacacgtgcttggtgcagctcagtgctgcctcggt 1266

Db 106230 TTGACTTTGAAACAGTTTAAACACAGTGTCTTGGTCAGCTCCAGTGTCTGTCGCCGTG 106171

QY 1267 cgggggtgagtggtgatcttgcctcttctgtctgttgatttgcagatgagtcgtg 1326

Db 106170 CGGGGTTGAGTGTGCATCTTTGCCCTTCTTCTGCTGTGATTTTGGCCAGATGATCTG 106111

QY 1327 cattatttgtacttttctatgtattataatcctgttagaagtcactaataaaggagtat 1386

Db 106110 CATTTATTGTACTTTTCTATGTATTATAATCCTGTAGAAGTCACCTAATAAAGGAGTA- 106050

QY 1387 tttttttgcagcttatcaatcagactgatctaagtgaagttaagtcacttaaaaa 1446

Db 106051 TTTTTTTGTCAGCTTATCAATCAATAGATGATCTAATGTGAATTAAGTATCCCTTAAAAA 105992

QY 1447 caaagcatctatttggcagaaaattgttcttctaaattcagtcatttgatctgtgag 1506

Db 105991 CAAAGCATCTATTTTGGCAGAAATGTGTTCTTAAATTCAGTCAATTGATATCTGTGAG 105932

QY 1507 acttcatttctcattcccttattgtcttttttagcaacataagaacaccatgagtcattt 1566

Db 105931 ACTTCAATATTTCTCATCCCTTTATTTGCTTTTGTAGCAACATAGAACCATGAGTCATTT 105872

QY 1567 tgcatttagagtatctgtataaatactcttgaaatactgaaatacaaaaggttaatgat 1626

Db 105871 TGTCAATTTAGAGTATCTGTATAAATCTTTGAAATACGTGAAATCAAAAGGTTAATGAT 105812

QY 1627 ttttgttcattctgattgttcatttattatctgttctcgttcgaagtgactaatttac 1686

Db 105811 TTTTGTTCATCTGATTTGTCATTTTATCTGTTATCGGTCTAAGTCTAATTAC 105752

QY 1687 ccatttgattttctgttagacagataaacttttaattttcnaatttggcagacactttt 1746

Db 105751 CCATTTGATTTTCTGTCTAGACAGATAACTTTTAATTTTCAAAATTTGGCAGACACTTTT 105692

QY 1747 tttttttttgaaactcttctccagatctgttcccaactgaacagccaccgtccc 1806

Db 105691 TTTTTTTTTTGAAATCTTTCTTCAGATCTGTGCCCCACTGAACAGCCACCCGTCCTC 105632

QY 1807 tcactgtcctggtgtccgattgggctggatggtgttggggcatgatgtgtgaggaactg 1866

Db 105631 TCACGTGCTGCTGCTCGATTTGGCTGGATGGTGTGGGCGCATGATGTGGAGAACTG 105572

QY 1867 gaagggtcttagtgcctgggttcagggtcgggcatctcttctgtgttgacacatcttttaa 1926

Db 105571 GAAGGTGCTTTAGTCTGTGGTTTCAGGGTCGGGCATTTCTTTGTTTGCACATCTTTTAA 105512

QY 1927 attttacacctttcttaagaattcttaataatgccgtcttaagttttttatataccaataatgctg 1986

Db 105511 ATTTTACACCTTTTCTTAAGAAATCTAATGCCGTCTTAAGTTTATATACCAATAATGCTG 105452

QY 1987 agctttaagtagtagatctggttagtcagacagacagtgatgatgga 2028

Db 105451 AGCTTAAAGTGTAGATCTGCTAGTACACAGACAGTGTGATGCA 105410

RESULT 7

AC025565 168012 bp DNA linear HTG 01-SEP-2000

LOCUS Homo sapiens chromosome 5 clone RP11-770E15, WORKING DRAFT

DEFINITION SEQUENCE, 16 unordered pieces.

ACCESSION AC025565

VERSION AC025565.3 GI:9960368

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 168012)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Sep 1, 2000 this sequence version replaced gi:7235344.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H_NH0770E15

----- Summary Statistics -----

Sequencing vector: M13; 100%

Chemistry: Dye-primer ET; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 160169 bases at least Q40

Consensus quality: 162474 bases at least Q30

Consensus quality: 164047 bases at least Q20

Insert size: 164000; agarose-fp

Insert size: 166512; sum-of-contigs

Quality coverage: 4.42 in Q20 bases; agarose-fp

Quality coverage: 4.39 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 8432: contig of 8432 bp in length

* 8433 8532: gap of unknown length

* 18030 18030: contig of 9498 bp in length

* 18031 18130: gap of unknown length

* 18131 26974: contig of 8844 bp in length

* 26975 27074: gap of unknown length

* 27075 44488: contig of 17414 bp in length

* 44489 44588: gap of unknown length

* 44589 56648: contig of 12060 bp in length

* 56649 56748: gap of unknown length

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 192994)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barne, N., Beckerly, R., Boguslavskiy, L., Boukhvalter, B.,
Brown, A., Castelle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollard, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-Oct-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 19, 2001 this sequence version replaced gi:7230119.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3524
Center clone name: 16_E18

----- Summary Statistics
Sequencing vector: M13; M7815; 52% of reads
Sequencing vector: Plasmid; n/a; 48% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181440 bases at least Q40
Consensus quality: 187416 bases at least Q30
Consensus quality: 189491 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 190394; sum-of-contigs
Quality coverage: 11.2 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 160053: contig of 160053 bp in length
* 160054 160153: gap of 100 bp
* 160154 160945: contig of 792 bp in length
* 160946 161045: gap of 100 bp
* 161046 161833: contig of 788 bp in length
* 161834 161933: gap of 100 bp
* 161934 162733: contig of 800 bp in length
* 162734 162833: gap of 100 bp
* 162834 163725: contig of 892 bp in length
* 163726 163825: gap of 100 bp
* 163826 164571: contig of 746 bp in length
* 164572 164671: gap of 100 bp
* 164672 165784: contig of 1113 bp in length
* 165785 165884: gap of 100 bp
* 165885 167095: contig of 1211 bp in length
* 167096 167195: gap of 100 bp
* 167196 167947: contig of 752 bp in length
* 167948 168047: gap of 100 bp
* 168048 169449: contig of 1402 bp in length
* 169450 169549: gap of 100 bp
* 169550 170722: contig of 1173 bp in length
* 170723 170822: gap of 100 bp

* 170823 171557: contig of 735 bp in length
* 171558 171657: gap of 100 bp
* 171658 172482: contig of 825 bp in length
* 172483 172582: gap of 100 bp
* 172583 173952: contig of 1370 bp in length
* 173953 174052: gap of 100 bp
* 174053 175031: contig of 979 bp in length
* 175032 175131: gap of 100 bp
* 175132 176068: contig of 937 bp in length
* 176069 176168: gap of 100 bp
* 176169 177688: contig of 1520 bp in length
* 177689 177788: gap of 100 bp
* 177789 177965: contig of 1277 bp in length
* 179066 179165: gap of 100 bp
* 179166 180432: contig of 1267 bp in length
* 180433 180532: gap of 100 bp
* 180533 181730: contig of 1198 bp in length
* 181731 181830: gap of 100 bp
* 181831 183033: contig of 1203 bp in length
* 183034 183133: gap of 100 bp
* 183134 184409: contig of 1276 bp in length
* 184410 184509: gap of 100 bp
* 184510 186294: contig of 1785 bp in length
* 186295 186394: gap of 100 bp
* 186395 188531: contig of 2137 bp in length
* 188532 188631: gap of 100 bp
* 188632 190451: contig of 1820 bp in length
* 190452 190551: gap of 100 bp
* 190552 192545: contig of 1994 bp in length
* 192546 192645: gap of 100 bp
* 192646 192994: contig of 349 bp in length.

FEATURES
Location/Qualifiers

source
1..192994
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_RPII="16E18"
/clone_lib="RPCI-11 Human Male BAC"

misc_feature

1..160053
/note="assembly_fragment"
clone_end:SP6
vector_side:left

misc_feature

160154..160945
/note="assembly_fragment"

misc_feature

161046..161833
/note="assembly_fragment"

misc_feature

161934..162733
/note="assembly_fragment"

misc_feature

162834..163725
/note="assembly_fragment"

misc_feature

163826..164571
/note="assembly_fragment"

misc_feature

164672..165784
/note="assembly_fragment"

misc_feature

165885..167095
/note="assembly_fragment"

misc_feature

167196..167947
/note="assembly_fragment"

misc_feature

168048..169449
/note="assembly_fragment"

misc_feature

169550..170722
/note="assembly_fragment"

misc_feature

170823..171557
/note="assembly_fragment"

misc_feature

171658..172482
/note="assembly_fragment"

misc_feature

172583..173952
/note="assembly_fragment"

misc_feature

174053..175031
/note="assembly_fragment"

misc_feature

175132..176068
/note="assembly_fragment"

misc_feature

176169..177688
/note="assembly_fragment"

```
misc_feature 17789..179065
/note="assembly_fragment"
misc_feature 179166..180432
/note="assembly_fragment"
misc_feature 180533..181730
/note="assembly_fragment"
misc_feature 181831..183033
/note="assembly_fragment"
misc_feature 183134..184409
/note="assembly_fragment"
misc_feature 184510..186294
/note="assembly_fragment"
misc_feature 186395..188531
/note="assembly_fragment"
misc_feature 188632..190451
/note="assembly_fragment"
misc_feature 190552..192545
/note="assembly_fragment"
misc_feature 192646..192994
/note="assembly_fragment"
vector_side:right"
clone_end:T7

BASE COUNT 59747 a 38890 c 36686 g 55068 t 2603 others
ORIGIN

Query Match      18.0%  Score 365.8;  DB 2;  Length 192994;
Best Local Similarity 69.2%  Pred. No. 1.2e-73;
Matches 529;  Conservative 0;  Mismatches 232;  Indels 4;  Gaps 2;

Qy 393  acccagagctcagctctccccagcacagccccagcttggctcaacgcgtagtaccag 452
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73613 ACCACAACACAGCAGCAACAAAGACCTCAGCCACAGCCCGCAGCTCACCAAGTGCAGCC 73554

Qy 453  ccctcagcagccacccagaccgctgggtggcccccagcaaacgagcggcgtttgg 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73553 TCAACAGCAGCAGCTGCAGATCCCTGGGTAGCTCCTCGTAAAGGGGAGCGAGCTTCAA 73494

Qy 513  gcagagcgaggggctggcagcagatagcaactctcctggaaacgtccagcctaattctgc 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73493 CCAGAACATGAGCGGCGAGTGAACACTTTGGTTAGGTGTTGACTGTCACTGCTTC 73434

Qy 573  cccc---agegtcgaatccccaccctccttgaataaactgaagctgctcacagctcaaa 629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73433 ACCTTCTAGTGTAGAAGTGATCCCGTCTGCTGGAAGAACTAAGGCCATAAACACTATA 73374

Qy 630  cccgaagaggttgagtggaatctgaaagcggcggtgtgttcataatcaagagactactc 689
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73373 TCCCAAGAGACTTGTATTGGAATCTGAAGAAATGAGCGTGTGTTTATAATTAAAGCTACTC 73314

Qy 690  tgagagcagacatccaccgctccattagtagtactccatctggtgtgagcagagcagcaaa 749
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73313 TGAGATGACATACATCGTTCCTAATAACTCTATCTGGTGTAGTACTGAGCATGGTAA 73254

Qy 750  caagcgctgtgacagcgtctccctgctgtagcagcaagggcgccgtctaccctgctctt 809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73253 TAAGCGTTTGGATGCAGCTTACCCTGCTCAATGGGAAAGGCCCACTCTATTACTCTT 73194

Qy 810  cagcgctcaatgggagtgggcatttttgggggtggcgagatgaagtcgcccgctggacta 869
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73193 CAGTGTGAATGGCAGTGSACATTTTGTGGAGTGGCTGAAATGAAGTCTGTGTGGACTA 73134

Qy 870  cggcaccagtgccgggtctgtgtctcaggaagtggaaggggaagtttgatgccagt 929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73133 TAATGCGTATGCTGGTGTCTCAGGATAAGTGGAGGGGCAAAATTTGAAGTTAAATG 73074

Qy 930  gatattgttaagatgtaccaccaataaccagctccggcacatcagggctggagataacga 989
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73073 GATCTTTGTCAAGATGTTCCCAATAACCAATTACGGCATATTTCGTTAGAAAATAATGA 73014

Qy 990  caacaaacccgttcacaaactccccgggacacccagagaggtgccccttagaaaaagccaagca 1049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73013 CAACAACCCGTTACCAATTTCAAGGGCAGACTCAAGAGAGTACCCCTAGAAAAAGCTAAGCA 72954
```

```
Qy 1050 agtctgaaattatcagttctacaaagcacacacaccttcacatcttcgacactttgtcca 1109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72953 AGTGCTTAAATAATTTGCTACTTTCAAGCATACCACTCAATCTTTGATGACTTTGCACA 72894
```

```
Qy 1110 ctacgagaagcg-ccagagaggaggagggtgtgctgcgaaggaagc 1153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72893 TTATGAAAAGCGTCAAGAAGAGGAGGAGGAGCCATCGTAGGGTAAAG 72849
```

RESULT 10

BC002559

LOCUS

DEFINITION Homo sapiens, high-glucose-regulated protein 8, clone MGC:739

IMAGE:3139250, mRNA, complete cds.

ACCESSION BC002559

VERSION BC002559.1 GI:12803468

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2309)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgrl.nih.gov

Shvchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tienson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 4 Row: 1 Column: 23

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 7705410.

Location/Qualifiers

1. .2309

/organism="Homo sapiens"

/db_xref="LocusID:51441"

/db_xref="taxon:9606"

/clone="MGC:739 IMAGE:3139250"

/tissue_type="Placenta, choriocarcinoma"

/clone_lib="NIH_MGC_21"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

372..2111

/codon_start=1

/product="high-glucose-regulated protein 8"

/protein_id="AAH02559.1"

/db_xref="GI:12803468"

/translation="MSASSLLEORPKGGKNGVSHQDGLNDLDDPEPVLSPQARP

NNAYTAMSDSYLPSYSPFSGLGEAAWSTGDTAMPYLTYSYGLSNFHPFLP

AMRGQPGALGTPFLGNGFNFFPSGIDFAWNGNDSTQSGYSNFTYAPSSL

GGAMIDGGSAPANETLKPCMTIDQMAALKIGSTEASNPVKVSGSITS

NIVASNSLPATIAPPKASWADIASKPAKQPKLTKNGTAGSSLPPIKHNNDIG


```
/codon_start=1
/product="high-glucose-regulated protein 8"
/db_xref="AA08813.1"
/translation="MSASSLLEQRPKGGCKGVKQNGSVHVKDGLNDDDDPEPVLSPQARP
NNATAMSDSYLPSYISGFSYSLGEAAWSTGGDTAMPYLTSTGQUSNPEHPFLPD
AMFQPGALGSTPFLGQHFNFPSGIDFSAMGNSSQGSQSSYSSNTAYAFSSL
GGAMIDGQSAFANETLNKAPGMTIDGMAALKLGSTEVASNPVKVGSVAGSGSITS
NIVASNSLPATAPKPAWADASKPAKQPKLTKNKTAGSLPPPTIKHNNDIG
TWNKGPVAKAPSAQLVQNTGQSPQVQOANNPPVAQASVGQOTOPLPPPP
OPALSVQOQAAQPTRWAPRNRSFGHNGVDNGVGSQAGSGSTPSEHPVLEKL
RSINYNPKDFDMLKHGRVFLIKSYSEDDIHRSIKYNWSTEGNRLDAAYFSMN
GKGPVYLLFSVNGSHFCGVAEMKSAVDYNTCAGVWSQDKWKRDFRWIPVKDPNS
QLRHIRENNENKPVNTRDQVEPVEKAKOVKLIASIKYKHTTSIFDDFSHYEKROE
RRKC"
BASE COUNT      775 a   612 c   630 g   713 t
ORIGIN
Query Match      16.8%; Score 340.6; DB 9; Length 2730;
Best Local Similarity 67.5%; Pred. No. 3.8e-68;
Matches 512; Conservative 0; Mismatches 239; Indels 8; Gaps 2;
QY 416 cacagccccagcttggctcaacgcagtcacagccctcagcagccaccaccacccc 475
  || || || || || || || || || || || || || || || || || || || ||
Db 1139 CTCACCTCCACACAGCCTGCCAGCTTTCAGTCCAGCAACAGGCGACTCAGCCAACCC 1198
  || || || || || || || || || || || || || || || || || || || ||
QY 476 gctgggtgccccacacacacacacacgcgcttgggcagacgagggcggtgagcgcg 535
  || || || || || || || || || || || || || || || || || || || ||
Db 1199 GCTGGGTAGCACCTCGGAACCCGCGAGTGGGTTCGGTCAATGAGGGGTGATGTAATG 1258
  || || || || || || || || || || || || || || || || || || || ||
QY 536 atagcaactctcctggaaacgctccagcctaatcttgcgccccagcgtcgaaatccccaccccg 595
  || || || || || || || || || || || || || || || || || || || ||
Db 1259 GAGTAGGACAGCTCAGGCGTGGTCTGGATCTACTCTCTC-----AGAACCCCAACCCAG 1312
  || || || || || || || || || || || || || || || || || || || ||
QY 596 tcttgaaaaaacaaagcgtgctcagctacacacccgaaagagtttgagtggaatctga 655
  || || || || || || || || || || || || || || || || || || || ||
Db 1313 TGTTCGAGAAGCTTCGGTCCATTAATACTATAACCCCAAGATTTTGACTGGAATCTGA 1372
  || || || || || || || || || || || || || || || || || || || ||
QY 656 aaagcggcggtgttcatcaacagagctactctgagacacacacccgctccatca 715
  || || || || || || || || || || || || || || || || || || || ||
Db 1373 ACATGCGCGGTTTTCATCATTAAGAGTACTCTGAGGACGATATTCACCGTTCGATTA 1432
  || || || || || || || || || || || || || || || || || || || ||
QY 716 agtactccatctggtgtgtagcacagagcagcgaacgagcgcctgacagcgcctccgcgt 775
  || || || || || || || || || || || || || || || || || || || ||
Db 1433 AGTATAAATTTTGGTGCACACAGAGCATGGTAACAAGAGACTGGATGCTGTATTCGTT 1492
  || || || || || || || || || || || || || || || || || || || ||
QY 776 gcatgagcagaaggggcccgtctacctgctctcagcgtcaatggagtgggcattttt 835
  || || || || || || || || || || || || || || || || || || || ||
Db 1493 CCATGAACGGGAAGGCCCGCTTACTTTTACTTTTTCAGTGTCAACGCGAGTGGACACTTCT 1552
  || || || || || || || || || || || || || || || || || || || ||
QY 836 gTggggTggcagatgaagtcgcccgtagactagcggcaccagtcgcccgggtctggtctc 895
  || || || || || || || || || || || || || || || || || || || ||
Db 1553 GTGGCGTGCCAGAAATGAATCTGCTGTGGATACACACATGTGCGAGGTGTGTGTGCC 1612
  || || || || || || || || || || || || || || || || || || || ||
QY 896 aggacaagtggaaagggaagtttgatgtccagtcgatttttggtaaggatgtaccacaata 955
  || || || || || || || || || || || || || || || || || || || ||
Db 1613 AGGACAAATGGAAGGTGCTTTTGTGTCAGGTGGATTTTGTGAAGGAGCGTTCCCAATA 1672
  || || || || || || || || || || || || || || || || || || || ||
QY 956 accagctccggcacatcaggctggagaaataacagacacacacacccggtccacaactcccggg 1015
  || || || || || || || || || || || || || || || || || || || ||
Db 1673 GCCAACTGGCACACATTCGCCCTAGAGAACACGAGAATAAACACGATGACCAACTTAGGG 1732
  || || || || || || || || || || || || || || || || || || || ||
QY 1016 acaccagaggtgccccttagaagaacgaacgaagcgaagcgtgctgaaatattcagttcctaca 1075
  || || || || || || || || || || || || || || || || || || || ||
Db 1733 ACACCTCAGGAAGTGCCTCTGGAAGAGCTAACAGGTGTGAAATATTATAGCCAGGTACA 1792
  || || || || || || || || || || || || || || || || || || || ||
QY 1076 agcacacacccctccatcttcgacgacttgcctacacagagagcgc--agagagggag 1133
  || || || || || || || || || || || || || || || || || || || ||
Db 1793 AGCACACCACTTCCATTTTGTGATGACTTCTCACACTATCAGAAACCCCAAGAGGAAGAA 1852
  || || || || || || || || || || || || || || || || || || || ||
QY 1134 gagggtggtcgcaaggaacggcagagtcgaaacaaaca 1172
  || || || || || || || || || || || || || || || || || || || ||
```

Db 1853 GAAAGCTGTTAAAGGACGCTCAAGGTCCTGGGAAATAA 1891

RESULT 14

BC014797

LOCUS BC014797 2079 bp mRNA linear ROD 04-OCT-2001

DEFINITION Mus musculus, clone MGC:11691 IMAGE:3962417, mRNA, complete cds.

ACCESSION BC014797

KEYWORDS BC014797.1 GI:15928653

SOURCE MGC.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2079)

Strausberg, R.

Direct Submission

Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalon@bcm.tmc.edu

Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 17 Row: k Column: 14

This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

Location/Qualifiers

1. 2079

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="MGC:11691 IMAGE:3962417"

/tissue_type="Mammary tumor. Metallothionlen-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."

/clone_lib="NCI_CGAP_Mam1"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

137..1876

/codon_start=1

/product="Unknown (protein for MGC:11691)"

/protein_id="AAH14797.1"

/db_xref="GI:15928654"

/translation="MSASSLLEQRPKGGCKGVKQNGSVHVKDGLNDDDDPEPVLSPQARP

NNATAMSDSYLPSYISGFSYSLGEAAWSTGGDTAMPYLTSTGQUSNPEHPFLPD

AMFQPGALGSTPFLGQHFNFPSGIDFSAMGNSSQGSQSSYSSNTAYAFSSL

GGAMIDGQSAFANETLNKAPGMTIDGMAALKLGSTEVASNPVKVGSVAGSGSITS

NIVASNSLPATAPKPAWADASKPAKQPKLTKNKTAGSLPPPTIKHNNDIG

TWNKGPVAKAPSAQLVQNTGQSPQVQOANNPPVAQASVGQOTOPLPPPP

OPALSVQOQAAQPTRWAPRNRSFGHNGVDNGVGSQAGSGSTPSEHPVLEKL

RSINYNPKDFDMLKHGRVFLIKSYSEDDIHRSIKYNWSTEGNRLDAAYFSMN

GKGPVYLLFSVNGSHFCGVAEMKSAVDYNTCAGVWSQDKWKRDFRWIPVKDPNS

QLRHIRENNENKPVNTRDQVEPVEKAKOVKLIASIKYKHTTSIFDDFSHYEKROE

EEYKKERQGRK"

BASE COUNT 595 a 505 c 499 g 480 t

ORIGIN

Query Match 16.8%; Score 340.2; DB 10; Length 2079;

Best Local Similarity 74.2%; Pred. No. 4.4e-68;
Matches 443; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

QY 583 gaatcccccctccttgaactgaagctgctcacagctacacccgaagaattt 642
Db 1286 GAGCCTCACCGGTGGGAAACTTCGGTCCATTAATACTAACCTTAAAGATTTC 1345

QY 643 gagggaatctgaagaacgagcgtgtgttcatcatcaagagctactctgaggaacatc 702
Db 1346 GACTGGAATCTGAACATGGCCGGTTCATCATTAAGAGCTACTCTGAGGACGATATC 1405

QY 703 caccgtccatttaagtactccatctggttagcagagacgagcgaacagcgcctggac 762
Db 1406 CACGGTTCATTAAGTATATCTGTGTGAGCAGACAGCATGTTGTAACAAGACTGGAT 1465

QY 763 agcccttcgctgctgagcagcaagggccgctctacctgctcttcacgctcaatggg 822
Db 1466 GCCCCTTATCGTTCATTAAGTGAAGGGTCCGCTGACTTCTTTCAGTGTCAACGGC 1525

QY 823 agtgggcatcttgggtggcgagatgaagtcccccgtggactcagggcaccagtgcc 882
Db 1526 AGTGGACACTTCTGTGAGTTGCAGAGATGAATCTGCTGGACTACAACACATGTGCA 1585

QY 883 ggggtctggtctcaggaacagtggaaggaagtttgatgtccagtgagtttggtaag 942
Db 1586 GGTGTGTGGTCCCAAGCAAAATGGAAGGGTGTTCGATGTTCAGATGGATTTTGTGAAG 1645

QY 943 gatgacccataaccagctccggcacatcagcgtgagataacgacacacaaacggctc 1002
Db 1646 GACGTTCCCAATAGCAACTCGGACATTCGCTAGAGAACACGAGATAAACCAAGT 1705

QY 1003 acaactccggggacacccagggagtgccctttagaaaaagccaaagcagtgctgaaatt 1062
Db 1706 ACCAACTCTAGGACACTCAGGAAGTGCCCTCGGAAAAGCTAAGCAGGTGTGAAATC 1765

QY 1063 atcagttctcagagcacacacacctccatcttcagcagcttgcactcaggaagcgc 1122
Db 1766 ATACCCAGCTACAGACACACACCTTCGATTTTGTGATGACTTCTCACACTATGAGAAACGC 1825

QY 1123 c-agagagagagagtggtgctgcaagaaagcgcagagctgcaaaacaaatgagg 1178
Db 1826 CAAGAGGAAGAAAGTGTGTTAAAAGGACGTCAGGTCGTGGGAAATAGAAAGCG 1882

RESULT 15
AC022766/c
LOCUS
DEFINITION Homo sapiens clone RP11-318E23, WORKING DRAFT SEQUENCE, 31
unordered pieces.
ACCESSION AC022766
VERSION AC022766.3 GI:8138013
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130984)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrera,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vasiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2000 this sequence version replaced gi:7712161.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5852
Center clone name: 318_E_23
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 118950 bases at least Q40
Consensus quality: 124428 bases at least Q30
Consensus quality: 126573 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 127984; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1066: contig of 1066 bp in length
* 1067 1166: gap of 100 bp
* 1167 2747: contig of 1581 bp in length
* 2748 2847: gap of 100 bp
* 2848 4584: contig of 1737 bp in length
* 4585 4684: gap of 100 bp
* 4685 6485: contig of 1805 bp in length
* 6490 6589: gap of 100 bp
* 6590 8380: contig of 1791 bp in length
* 8381 8480: gap of 100 bp
* 8481 10218: contig of 1738 bp in length
* 10219 10318: gap of 100 bp
* 10319 12338: contig of 2020 bp in length
* 12339 12438: gap of 100 bp
* 12439 14278: contig of 1840 bp in length
* 14279 14378: gap of 100 bp
* 14379 16225: contig of 1847 bp in length
* 16226 16325: gap of 100 bp
* 16326 18627: contig of 2302 bp in length
* 18628 18727: gap of 100 bp
* 18728 21450: contig of 2723 bp in length
* 21451 21550: gap of 100 bp
* 21551 25430: contig of 3880 bp in length
* 25431 25530: gap of 100 bp
* 25531 28870: contig of 3340 bp in length
* 28871 28970: gap of 100 bp
* 28971 32940: contig of 3970 bp in length
* 32941 33040: gap of 100 bp
* 33041 35975: contig of 2935 bp in length
* 35976 36075: gap of 100 bp
* 36076 39460: contig of 3385 bp in length
* 39461 39560: gap of 100 bp
* 39561 43197: contig of 3637 bp in length

```
* 43198 43297: gap of 100 bp
* 43298 46333: contig of 3036 bp in length
* 46334 46433: gap of 100 bp
* 46434 51625: contig of 5192 bp in length
* 51626 51725: gap of 100 bp
* 51726 55173: contig of 3448 bp in length
* 55174 55273: gap of 100 bp
* 55274 59400: contig of 4127 bp in length
* 59401 59500: gap of 100 bp
* 59501 64807: contig of 5307 bp in length
* 64808 64907: gap of 100 bp
* 64908 68668: contig of 3761 bp in length
* 68669 68768: gap of 100 bp
* 68769 72524: contig of 3756 bp in length
* 72525 72624: gap of 100 bp
* 72625 77709: contig of 5085 bp in length
* 77710 77809: gap of 100 bp
* 77810 83970: contig of 6161 bp in length
* 83971 84070: gap of 100 bp
* 84071 89106: contig of 5036 bp in length
* 89107 89206: gap of 100 bp
* 89207 96647: contig of 7441 bp in length
* 96648 96747: gap of 100 bp
* 96748 105147: contig of 8400 bp in length
* 105148 105247: gap of 100 bp
* 105248 118840: contig of 13593 bp in length
* 118841 118940: gap of 100 bp
* 118941 130984: contig of 12044 bp in length.
```

FEATURES

source

```
1..130984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-318E23"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1..1066
/note="assembly_fragment"
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
/note="assembly_fragment"
55274..59400
/note="assembly_fragment"
59501..64807
/note="assembly_fragment"
64908..68668
/note="assembly_fragment"
68769..72524
/note="assembly_fragment"
clone_end:sp6
vector_side:right"
72625..77709
/note="assembly_fragment"
77810..83970
/note="assembly_fragment"
84071..89106
/note="assembly_fragment"
clone_end:T7
vector_side:right"
89207..96647
/note="assembly_fragment"
96748..105147
/note="assembly_fragment"
105248..118840
/note="assembly_fragment"
118941..130984
/note="assembly_fragment"
```

Query Match

16.4%; Score 333.4; DB 2; Length 130984;

Best Local Similarity 68.0%; Pred. No. 3.5e-66;

Matches 482; Conservative 0; Mismatches 221; Indels 6; Gaps 1;

QY 416 cacagccccagcttggctcaacgcagtcacagagcctcagagccctcagcagccacccagacc 475
DB 82013 CTCACCTCCACACAGCGCTGCCAGCTTTCAGTCAGCAACAGCAGCTCAGCCACCC 81954

QY 476 gctgggttggccacgcaacgaaacgcgcgtttgggcagagcgaggggctggcagcg 535
DB 81953 GCTGGGTAGCACCTCGGAACCGTGGCAGTGGTTCGGTCATAATGGGTGGTAAATG 81894

QY 536 atagcaactctctggaaacgtccagcctaattctgtcccccagcgctcgaatcccccgcg 595
DB 81893 GAGTAGGACAGTCTCAGGCTGGTTCTGGATCTACTCTTC-----AGAACCCACCCAG 81840

QY 596 tcttgaaaaactgaaggctgctcacagctacaacccgaaagagtttggagtgaaactga 655
DB 81839 TCTTGAGAGAGCTTCGGTCCATTAACTAATAACCCCAAGATTTTGACTGAATCTGA 81780

QY 656 aaagcggcggtgttcatcatcaagagctactctgagagcagacatccaccgctccatca 715
DB 81779 AACATGGCGGGTTCATCATTAAGAGCTACTCTGAGGACGATATTCACCGTTCATTTA 81720

QY 716 agtactccatctggtgtagcacagagcaacgaagcgctggacagcgcttcctcgct 775
DB 81719 AGTATAAATATTGGTGCAGCAGCAGCATGGTAACAAGAGACTGGATGCTGTATTATCGTT 81660

QY 776 gcatgagcagcaaggcgccgctctacctgctctcagcgtcaatgagtgaggcattttt 835
DB 81659 CCATGAACGGGAAGGCCCGCTTTACTTACTTTTCAGTGTCAACGSCAGTGGACACTTCT 81600

QY 836 gtggggtggccagatgaagtcccccgtggactcggcaccagtgccgggggtctggtctc 895
DB 81599 GTGGCGTGGCAGAAATGAAATTCGCTGTGTGACTACAAACATGTGCAGGTGTGTGTC 81540

QY 896 aggacaagtggagggaagtttgatgtccagtggttttttaaggtatgaccacaata 955
DB 81539 AGGACAAATGGAAGGGTCCGTTTGTGATGTCAGGTGGATTTTGTGAAGGACGTTCCCAATA 81480

QY 956 accagctccggcacatcaggctggagaataacgacaacaaacgggtcacaaaactccggg 1015
DB 81479 GCCAACTGGACACATTCGCCTAGAGAACACAGAGANAATAACCACTGACCACTCTAGGG 81420

QY 1016 acaccagagggtgcctcttagaaaaagcccaagcgaagtgtgaaaaattatcagttcctaca 1075

```
||||| ||||| ||||| | ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 81419 ACACTCAGGAAGTGCCTCTGGAAAAGGCTAAGCAGGTGTTGAAAATTATAGCCACCTACA 81360
Qy 1076 agcacacaacctccatcttcgacgactttgtcactacgagaagcgcca 1124
||||| || ||||| || ||||| . ||||| ||||| |||||
Db 81359 AGCACACCACTTCCATTTTGTGACTTCTCACACTATGAGAAAGGCCA 81311
```

Search completed: July 29, 2002, 18:09:38
Job time: 11004 sec

THIS PAGE BLANK (USPTO)